BAYESIAN APPLICATIONS
IN ECONOMETRICS

A thesis presented for the degree of
Doctor of Philosophy
in Economics
at
the University of Canterbury
Christchurch, New Zealand

by

D.E.A. Giles

February 1975

CONTENTS

CHAPTER		PAGE
	ABSTRACT	1
I.	INTRODUCTION	3
	I. Introductory comments	3
	(1) General background	3
	(2) Bayesian methods in econometrics .	6
	II. Outline of the thesis	9
II.	BASIC RESULTS	12
	I. Definitions	12
	II. Bayes' theorem	13
	III. Estimation	15
	IV. Prediction	17
	V. Model comparisons	18
III.	A M.S.E. COMPARISON OF TWO M.E.L. ESTIMATORS FOR THE MULTIPLE REGRESSION MODEL	21
	I. Introduction	21
	II. Preliminary results	23
	III. Comparing $\hat{\beta}$ and $\tilde{\beta}$	26
	IV. Further analytic results	32
	(1) Specification of $\bar{\beta}$	33
	(2) Specification of A	37
	(3) Multicollinearity and precision . .	48
	V. A special prior p.d.f.	51
	VI. A test statistic	53
	(1) A formal test	53
	(2) An ad hoc test	59
	VII. Conclusions	80

CHAPTER PAGE

IV. BAYESIAN SEASONAL ADJUSTMENT OF ECONOMIC TIME SERIES 81

I. Introduction 81

II. Explaining seasonality 82

III. Statistical properties 83

IV. The systematic component 85

V. Estimating the seasonal influence . . 87

VI. Model comparisons 90

VII. Regression relationships with seasonal data 92

VIII. Conclusions 94

V. BAYESIAN ANALYSIS OF DISTRIBUTED LAG MODELS: A SURVEY 95

I. Introduction 95

II. Infinite lag models 97

(1) Rational lags 97

(2) Other contributions 100

III. Finite lag models 101

IV. Conclusions 104

VI. DISCRIMINATING BETWEEN AUTOREGRESSIVE FORMS: A MONTE CARLO COMPARISON OF BAYESIAN AND AD HOC METHODS 106

I. Introduction 106

II. B.P.O. Analysis 111

III. The Monte Carlo study 114

(1) M_1 is true 115

(2) M_2 is true 117

(3) Limitations of the tests 117

CHAPTER PAGE

IV. The choice of data and parameters . . 119

(1) Exogenous variable 119

(2) Parameter values 119

(3) Random disturbances 120

(4) Initial value of y 121

V. Results 121

(1) M_1 is true 122

(2) M_2 is true 128

(3) Summary 133

VI. Conclusions 135

VII. BAYESIAN INFERENCE AND THE RESTRICTED ALMON ESTIMATOR. PART I: THEORETICAL RESULTS 137

I. Introduction 137

II. The classical P.A.E. 138

(1) Estimation 138

(2) Linear restrictions 139

(3) A Critique 140

III. Bayesian considerations 143

IV. Re-parameterizing the model 147

V. Bayesian estimation 151

(1) Serially independent errors . . . 151

(2) Serially correlated errors . . . 156

VI. Random L 159

(1) Serially independent errors . . . 160

(2) Serially correlated errors . . . 163

VII. Model comparisons 165

(1) Computation of the B.P.O. 165

(2) A Qualification 169

VIII. Conclusions 170

CHAPTER PAGE

VIII. BAYESIAN INFERENCE AND THE RESTRICTED ALMON ESTIMATOR. PART II: AN APPLICATION TO CURRENT PAYMENTS FOR NEW ZEALAND IMPORTS . . 172

I. Introduction 172

II. Data 173

III. Seasonal adjustment 175

IV. Estimation 177

V. Prior knowledge 179

(1) The model space 180

(2) The parameter space 180

VI. Posterior distributions 183

(1) L fixed and known; errors serially independent 184

(2) L fixed and known; errors serially correlated 185

(3) L random; errors serially independent 187

(4) L random; errors serially correlated 189

VII. B.P.O. and point predictions 191

(1) B.P.O. analysis 191

(2) Point predictions 192

VIII. Results 193

(1) Discussion of tabulated results 193

(2) Diagrammatic results 221

(3) Appraisal of the Bayes-Almon estimator 232

(4) Economic implications 236

IX. Concluding remark 237

CHAPTER		PAGE
IX.	SUMMARY AND CONCLUSIONS	238
	ACKNOWLEDGEMENTS	242
	REFERENCES	244
	APPENDIX I: SOME USEFUL RESULTS	253
	APPENDIX II: SOME RESULTS WITH THE ONE-REGRESSOR MODEL	258
	APPENDIX III: LAG DISTRIBUTION SHAPES	260
	APPENDIX IV: RAW DATA FOR CHAPTER VIII . . .	264
	APPENDIX V: SEASONALLY ADJUSTED DATA . . .	267

LIST OF TABLES

TABLE		PAGE
III.VI.1	Hybrid R^2 values	68
III.VI.2	$\beta = 1.0$; $\sigma^2 = 1.0$; $n = 10$	71
III.VI.3	$\beta = 1.0$; $\sigma^2 = 1.0$; $n = 30$	71
III.VI.4	$\beta = 2.0$; $\sigma^2 = 1.0$; $n = 10$	72
III.VI.5	$\beta = 2.0$; $\sigma^2 = 1.0$; $n = 30$	72
III.VI.6	$\beta = 3.0$; $\sigma^2 = 1.0$; $n = 10$	73
III.VI.7	$\beta = 3.0$; $\sigma^2 = 1.0$; $n = 30$	73
III.VI.8	$\beta = 1.0$; $\sigma^2 = 2.0$; $n = 10$	74
III.VI.9	$\beta = 1.0$; $\sigma^2 = 2.0$; $n = 30$	74
III.VI.10	$\beta = 2.0$; $\sigma^2 = 2.0$; $n = 10$	75
III.VI.11	$\beta = 2.0$; $\sigma^2 = 2.0$; $n = 30$	75
III.VI.12	$\beta = 3.0$; $\sigma^2 = 2.0$; $n = 10$	76
III.VI.13	$\beta = 3.0$; $\sigma^2 = 2.0$; $n = 30$	76
III.VI.14	Estimated power functions	77
VI.IV.1	Hybrid R^2 values	121
VI.V.1	M_1 true; $\gamma = 0.90$; $\rho = 0.30$	124
VI.V.2	M_1 true; $\gamma = 0.90$; $\rho = 0.60$	125
VI.V.3	M_1 true; $\gamma = 0.90$; $\rho = 0.90$	126
VI.V.4	$G \leq 5\%$	127
VI.V.5	M_2 true; $\beta = 0.09$; $\lambda = 0.90$	130
VI.V.6	M_2 true; $\beta = 0.45$; $\lambda = 0.50$	131
VI.V.7	M_2 true; $\beta = 0.81$; $\lambda = 0.10$	132
VI.V.8	$\frac{1}{2}[\overline{p(\mathrm{I})} + \overline{p(\mathrm{II})}]$	134
VIII.VIII.1	Part 1: diffuse prior	194
VIII.VIII.2	Part 1: diffuse prior	195
VIII.VIII.3	Part 1: diffuse prior	196
VIII.VIII.4	Part 1: loose prior	197

TABLE		PAGE
VIII.VIII.5	Part 1: loose prior	198
VIII.VIII.6	Part 1: loose prior	199
VIII.VIII.7	Part 1: tight prior	200
VIII.VIII.8	Part 1: tight prior	201
VIII.VIII.9	Part 1: tight prior	202
VIII.VIII.10	Part 2: diffuse prior	203
VIII.VIII.11	Part 2: diffuse prior	204
VIII.VIII.12	Part 2: diffuse prior	205
VIII.VIII.13	Part 2: loose prior	206
VIII.VIII.14	Part 2: loose prior	207
VIII.VIII.15	Part 2: loose prior	208
VIII.VIII.16	Part 2: tight prior	209
VIII.VIII.17	Part 2: tight prior	210
VIII.VIII.18	Part 2: tight prior	211
VIII.VIII.19	Part 3	212
VIII.VIII.20	Part 3	213
VIII.VIII.21	Part 3	214
VIII.VIII.22	Part 4	215
VIII.VIII.23	Part 4	216
VIII.VIII.24	Part 4	217

LIST OF FIGURES

FIGURE		PAGE
III.IV.1	Effects of varying θ	34
III.IV.2	λ as a function of θ	36
III.IV.3	An extreme curve décolletage	42
III.IV.4	Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^* > \sigma^2$	44
III.IV.5	Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^* = \sigma^2$	44
III.IV.6	Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^* < \sigma^2$	44
III.IV.7	λ as a function of ϕ	47
III.V.1	The curve décolletage when $A \propto (X'X)$.	53
III.VI.1	Estimated power curves	78
VIII.VIII.1	Part 1, M_2 (diffuse prior)	222
VIII.VIII.2	Part 1, M_2 (loose prior)	222
VIII.VIII.3	Part 1, M_2 (tight prior)	222
VIII.VIII.4	Part 1, M_2 (posteriors)	223
VIII.VIII.5	Part 1, M_2 (joint posteriors)	223
VIII.VIII.6	Part 2, M_3 (diffuse prior)	224
VIII.VIII.7	Part 2, M_3 (loose prior)	224
VIII.VIII.8	Part 2, M_3 (tight prior)	225
VIII.VIII.9	Part 2, M_3 (comparative posteriors). .	225
VIII.VIII.10	Part 3, S_1 (diffuse prior)	226
VIII.VIII.11	Part 3, S_1 (loose prior)	226
VIII.VIII.12	Part 3, S_1 (tight prior)	227
VIII.VIII.13	Part 3, S_1 (comparative posteriors). .	227
VIII.VIII.14	Part 4, S_1 (diffuse prior)	228
VIII.VIII.15	Part 4, S_1 (loose prior)	229
VIII.VIII.16	Part 4, S_1 (tight prior)	230

FIGURE		PAGE
VIII.VIII.17	Part 4, S_1 (posteriors)	231
A.1	$P = 1$	260
A.2	$P \geq 2$	260
A.3	$P \geq 2$	261
A.4	$P \geq 2$	261
A.5	$P \geq 3$	261
A.6	$P \geq 3$	261
A.7	$P \geq 3$	261
A.8	$P \geq 4$	261
A.9	Seasonally adjusted current payments .	267
A.10	Seasonally adjusted c.i.f. imports .	268

ABSTRACT

The thesis considers several related aspects of Bayesian inference in econometrics. Particular attention is given to model-comparisons, distributed lags, and the sampling properties of estimators.

In Chapter III the natural-conjugate Bayes ($\tilde{\beta}$) and Ordinary Least Squares ($\hat{\beta}$) estimators for the linear model are compared, and a condition is derived and investigated under which $\tilde{\beta}$ is preferred to $\hat{\beta}$ in terms of matrix mean squared error. In a limiting case a test statistic is obtained and shown to be related to another well-known test. Two observable substitute statistics are shown to be consistent but upward-biased. The bias is studied in a limited Monte Carlo experiment.

Bayesian inferential methods are advocated in Chapter IV for the seasonal adjustment of economic time-series. This is motivated by Chapter III and the application in Chapter VIII. A well-known classical procedure is shown to be a special case of the Bayesian method.

Bayesian analyses of distributed lag models are surveyed in Chapter V, and Chapter VI considers the problem of discriminating between a Koyck distributed lag model and a regression model with autocorrelated disturbances. A Monte Carlo study compares several interpretations of an *ad hoc* rule-of-thumb proposed by Griliches with Bayesian Posterior Odds analysis, and the latter is found to be generally superior to the former.

Chapters VII and VIII present a Bayesian interpretation of the Almon estimator, treating theoretical results and an application to some New Zealand data. The theory generalizes that of Zellner and Williams, special attention being paid to: prior information; unknown lags; model comparisons; and autocorrelation. Although the methodological problems associated with the classical Almon estimator are overcome, the computational cost is increased substantially.

The evidence presented suggests that several econometric problems may be handled more satisfactorily by Bayesian methods than by classical methods.

CHAPTER I

INTRODUCTION

I. INTRODUCTORY COMMENTS

(1) General Background

In recent years the theory of mathematical statistics has been liberated to some extent from the firm grip of the "frequentist" (or "classical") school associated with Sir Ronald Fisher in the United Kingdom, and with J. Neyman and E.S. Pearson in the United States.[1] There has been renewed interest in an old principle first suggested by Thomas Bayes in 1763.

Although Bayes is well known for the theorem which bears his name, in fact it seems that he has even closer links with the modern "Bayesian" school, in that he shared their concept of probability.[2] Modern Bayesians differ from the frequentists by adopting a subjectivistic (personalistic) view of probability, rather than defining a probability as a long-run frequency. Although the use of Bayes' Theorem is important in modern Bayesian analysis, it is also used (correctly) in various contexts by frequentists.

The Bayesian approach draws on the contributions of de Finetti (1937), (1972); Ramsey (1950); Savage (1954), (1961); and Jeffreys (1961). This subjectivistic interpretation makes it meaningful to attach probabilities to once-and-for all events, parameters, hypotheses or models.

---

1. See Bartlett (1965); Plackett (1966); Lindley (1965).
2. Plackett, op. cit., pp.252-253.

The value of a subjective probability represents a personal degree of belief, based on the individual's present state of knowledge. As data are observed, this state of knowledge changes, and the *a priori* degrees of belief must be updated to the *a posteriori* state. This updating is effected through Bayes' Theorem, thus introducing the basic principle of "learning from experience". The process may continue sequentially if additional data become available. The previous *a posterior* state then becomes the new *a priori* state, and Bayes' Theorem is applied again, *etc*..

Expressing this somewhat differently, *a priori* information is combined with sample information, the total amount being summarized in the resulting *a posteriori* distribution. This distribution contains all of the available information. However, Bayesians often wish to make decisions on the basis of the posterior distribution[3], and at this stage the loss structure of actions is taken into account explicitly. Being based on the earlier work of von Neumann and Morgenstern (1947) and Wald (1950), Bayesian decision theory pursues the principle of acting to maximize expected utility.

The Bayesian philosophy should be (and has been) scrutinized on at least two fundamental levels. First, the concept of subjective probability itself[4] has prompted a heated debate. An individual's (subjective) probability of the outcome of an event is determined by the least odds at which he is prepared to bet on that outcome, such

---

3. For example, one may wish to draw statistical inferences, such as choosing a point estimate.
4. For example, see Kyburg and Smokler (1964),pp.3-15.

bets being measured in terms of utility. In all but the simplest of situations the introspection required to determine the prior distribution precisely would be enourmous.[5]

Secondly, there is the explicit introduction of a model of "learning from experience", arising with the use of Bayes' Theorem. This view of the learning process is in direct conflict with that advocated by Popper (1965), Kuhn (1962) and Feyerabend (1965), for example. In their view, we learn from our mistakes, and progress as a result of criticism and by trying to refute earlier propositions.[6]

The philosophical debate surrounding Bayes' Theorem is as old as the theorem itself, and the controversy over the subjective view of probability is by no means resolved either. These issues are of fundamental importance and will undoubtedly continue to pose problems for philosophers of mathematics for some time.

However, such matters are not the subject of this thesis, and are raised here only to provide a setting for what is to follow. Despite the controversy which still surrounds the Bayesian school of statistical thought, its adherents have made many important contributions in recent years, and its acceptance appears to be spreading.[7] Thus,

---

5. See Plackett, op.cit., pp.252-253; Raiffa and Schlaifer (1961), pp.59-62.
6. For example, see Rothenberg (1969), pp.200-204.
7. To some extent this is reflected in the appearance of texts on mathematical statistics which are written primarily from a Bayesian viewpoint. For example, de Groot (1970), and Box and Tiao (1973).

for the purposes of this thesis we accept the meaningfulness and usefulness of the Bayesian philosophy, at least in certain situations, and consider some aspects of its use in econometric theory.

(2) Bayesian Methods in Econometrics

The recent developments in mathematical statistics have been reflected in several of the disciplines based on this theory, one of these being econometrics. By virtue of its historical development, econometrics is firmly based on the principles of frequentist theory, though lately there has been increasing interest in the use of Bayesian methods for analysing econometric problems.

The supporters of a Bayesian approach in econometrics, notably Arnold Zellner, have argued convincingly for its wider adoption.[8] If there is one principal self-supporting argument advanced by these Bayesians it is that their approach is based on a simple and unified set of principles which provides the means for inference and decision-making in a wide variety of situations. According to its adherents, it is in this respect that the Bayesian approach in econometrics stands above its non-Bayesian rivals, the latter being seen as a collection of "ad hockeries".

The specific advantages of the Bayesian approach to econometric analysis are essentially those that apply at a more general level in mathematical statistics. These advantages have been summarized well by Zellner (1969), and

8. For example, see Zellner (1969). Much of Zellner's earlier work in the area of Bayesian inference in econometrics is synthesized in Zellner (1971).

relate to the main areas of econometric theory: estimation; hypothesis testing; prediction; control; the incorporation of prior information; the handling of "nuisance" parameters; and specification analysis.

Of prime concern in econometrics[9] are the two major areas of inference: estimation, and hypothesis testing. In the former, the main contribution of the Bayesian approach is that unknown parameters are treated as random variables and are assigned prior probability density functions (p.d.f.'s) or probability mass functions (p.m.f.'s). Sample information is introduced via the likelihood function, and then posterior p.d.f.'s (or p.m.f.'s) are produced by means of Bayes' Theorem. Once the loss structure is made specific, the posterior distributions form the basis for point or interval estimation[10], if such estimates are desired. For example, see Tiao and Zellner (1964).

With regard to the latter, the subjectivistic interpretation of probability makes it meaningful to compare non-nested models or hypotheses. In general, such comparisons cannot be formalized within the frequentist framework, though unfortunately ad hoc attempts are often made. Major contributions to the Bayesian theory of model comparisons have been made by Thornber (1966), Geisel (1970), Lempers (1971), and Gaver (1974). Bayesian and non-Bayesian methods in this area have been surveyed

---

9. Again, the comments here in general apply to mathematical statistics, not just to econometrics.
10. In contrast, classical theory treats the parameters as fixed unknowns, and, in general, inferences are based solely on sample information.

by Gaver and Geisel (1974). Again, the analysis relates to any parametric statistical models, but the presentation by these authors centres on economic problems, and economic applications are given.

There are many troublesome areas of traditional econometric theory which may be handled formally and efficiently by Bayesian methods. Examples are the problems of model comparisons; the handling of "nuisance" parameters[11]; specification analysis[12]; and distributed lag models[13], to name a few. As might be expected, however, these gains are not without cost.

For example, although it is helpful to have a formal means of using prior knowledge, it may be very difficult to formulate a satisfactory prior p.d.f. (or p.m.f.) in practice. In such cases a Bayesian might suggest trying a variety of prior distributions (including one devoid of informational content[14]) in order to test the sensitivity of inferences to the choice of prior p.d.f..

Further, in practice Bayesian methods in econometrics (and elsewhere) may have to be limited to rather simple[15] models, unless the prior p.d.f. can be chosen in such a way that it can be combined with the likelihood function

---

11. See Zellner (1971), pp.21-22.
12. See Lempers, op.cit., pp.47-65.
13. See Chapter V of this thesis.
14. For a discussion of prior ignorance, see Zellner (1971), pp.41-53, or Box and Tiao, op.cit., pp.25-60.
15. Here a "simple" model is one involving few unknown parameters.

analytically.[16] In general, numerical approximations must be used when integrating to normalize the posterior p.d.f. and to obtain its moments. In such cases the computational cost of even simple techniques as Simpson's rule is prohibitive in parameter spaces of even moderate dimension.[17]

However, as already noted, a wide range of important econometric problems which cause difficulties in a frequentist setting, are well suited to Bayesian analysis. Despite conceptual and practical difficulties, the Bayesian framework has much to offer when analysing problems of this type, and considerable progress has been made in a variety of areas.

A belief in this (limited) usefulness of Bayesian inference in econometrics underlies this thesis.

II. OUTLINE OF THE THESIS

The thesis considers a number of related problems concerned with the use of Bayesian methods of inference in econometrics. The layout is as follows. Chapter II contains some basic notation and results related to Bayesian estimation, hypothesis testing and prediction.

---

16. Two such types of prior p.d.f. are Jeffreys' "diffuse" prior, and the Natural-Conjugate priors suggested by Raiffa and Schlaifer, op.cit., pp.43-76.
17. Very roughly, the processing time when applying Simpson's rule is proportional to m^k, where m is the number of intervals used in the approximation, and k is the dimension of the parameter space.

This material is drawn primarily from Zellner (1971; Ch. 2, 3), and from Geisel (1970; Ch. 1, 2), and is used repeatedly throughout the thesis.

The rest of the thesis falls broadly into two sections. Chapter III is quasi-Bayesian in philosophy, lying between the frequentist and Bayesian schools, in that it considers the sampling properties of certain Bayes estimators. Using general[18] Mean Squared Error (M.S.E.) as a basis, two estimators for the linear regression model are compared in various situations. This M.S.E. basis forms part of the motivation in Chapter IV for a Bayesian interpretation of the problem of seasonally adjusting an economic time-series. The approach used here amounts to a particular application of the usual Bayesian methods of inference in parametric models.

The second, and major section of the thesis centres on the Bayesian analysis of distributed lag models. Such models have been analysed already by different Bayesian methods under various conditions, the relevant literature being surveyed in Chapter V. One of the main reasons for concentrating on this type of model here is that it raises a wide variety of econometric problems which are very difficult to solve by classical methods, but which may be handled more successfully by adopting a Bayesian approach.

Chapter VI analyses a particular problem of model discrimination which arises with certain infinite distributed lag models, and compares Bayesian Posterior Odds (B.P.O.)

---

18. See Goldberger (1964), p.129.

analysis with an *ad hoc* approach.

The Polynomial Approximation (Almon) estimator for finite distributed lag models has received a lot of attention in the literature since its conception, partly as a result of the methodological problems which it poses for frequentists. A general Bayesian interpretation of this estimator is developed in Chapter VII, this being an extension of the earlier contribution by Zellner and Williams (1973). This estimator is applied to some New Zealand data in Chapter VIII. Some concluding remarks appear in Chapter IX.

CHAPTER II

BASIC RESULTS

Here, and throughout the thesis the symbol p denotes any general prior or posterior p.d.f. or p.m.f., regardless of mathematical form. It may relate to parameters, observations, or models, but the meaning will be clear from the context in which it is used.[1]

I. DEFINITIONS

We define four spaces required for a discussion of Bayesian decision theory:

The sample space, $Y = \{y\}$, is the set of all possible values of the sample observations for the experiment of interest.

The action space, $A = \{a\}$, is the set of possible actions related to a particular decision problem.

The model space, $\mathcal{M} = \{M\}$, is the set of possible models or hypotheses for the experiment being considered.[2]

The parameter space, $\Omega_i = \{\theta_i\}$, for the ith. model in $\mathcal{M}$ is the set of unknown parameters associated with that model.

Both $\mathcal{M}$ and the Ω_i's are state spaces in the usual terminology. Further, a parametric statistical model of a

---

1. Detailed descriptions of the basic Bayesian framework are given by Raiffa and Schlaifer, op.cit., de Groot, op.cit., and Box and Tiao, op.cit..
2. This space is restricted to be finite or countable, for reasons described below. However, this restriction can be relaxed in general.

stochastic process is a family of data densities, dependent on a (usually finite) number of parameters and a specified set of predetermined variables, and a prior density for the parameters of the data densities.

For each model in $\mathcal{M}$, the data density is given by $p(y|\theta_i,M_i)$, which when viewed as a function of the unknown parameters is the likelihood function, $\ell(\theta_i|y,M_i)$. The prior p.d.f. (or p.m.f.) for the parameters of M_i is given by $p(\theta_i|M_i)$. Finally, a prior p.m.f. is defined over the elements of $\mathcal{M}$, so that the prior mass of the ith. model is $p(M_i)$.

II. BAYES' THEOREM

The conditional (on the model) data density is obtained as

$$p(y|M_i) = \int_{\Omega_i} p(y|\theta_i,M_i)p(\theta_i|M_i)d\theta_i, \qquad \text{II.II.1}$$

and the marginal data density is

$$p(y) = \sum_i p(y|M_i)p(M_i) \qquad \text{II.II.2}$$

Then, applying Bayes' Theorem, the (joint) posterior p.d.f. for the parameters of the ith. model is

$$p(\theta_i|y,M_i) = \{p(\theta_i|M_i)p(y|\theta_i,M_i)\}/p(y|M_i) \qquad \text{II.II.3}$$

Alternatively, this may be expressed as

$$p(\theta_i|y,M_i) \propto p(\theta_i|M_i)\ell(\theta_i|y,M_i)$$

where the proportionality constant is[3]

$$\{\int_{\Omega_i} p(\theta_i|M_i)\ell(\theta_i|y,M_i)d\theta_i\}^{-1}$$

Further, the posterior probability of the ith. model is given by

$$p(M_i|y) = \{p(M_i)p(y|M_i)\}/p(y), \qquad \text{II.II.4}$$

or,

$$p(M_i|y) \propto p(M_i)p(y|M_i)$$

where the proportionality constant is $\{\sum_i p(M_i)p(y|M_i)\}^{-1}$.

In many instances, attention centres on one (or a subset) of the parameters in the vector θ_i. In such cases, marginal posterior p.d.f.'s for these parameters may be derived. Partition θ_i and Ω_i such that

$$\theta_i' = (\theta_{1i}', \theta_{2i}'),$$

$$\Omega_i = (\Omega_{1i} \cup \Omega_{2i}).$$

Then,

$$p(\theta_{1i}|y) = \int_{\Omega_{2i}} p(\theta_{1i}, \theta_{2i}|y)d\theta_{2i} \qquad \text{II.II.5}$$

---

3. Here and throughout the thesis the notation is simplified by using the single symbol $\int$ even for multi-dimensional integrals.

where $\theta_{1i} \in \Omega_{1i}$; and a symmetric result holds if $p(\theta_{2i}|y)$ is required. Such integration is also useful for eliminating "nuisance" parameters, once their influence on the posterior p.d.f. for the remaining parameters has been accounted for.

III. ESTIMATION

We define a decision rule, d, as a function mapping Y to A. Further, a loss function L is a real function describing the loss of utility (in the von Neumann-Morgenstern sense) resulting when action a is taken for a particular true state of nature. This state of nature will be a point in Ω_i here, but in Section V below it will be a model in $\mathcal{M}$. Thus, here L maps $(\Omega_i \times A)$ to $[0,\infty)$. The notation $L(\theta_i,\hat{\theta}_i)$ denotes the loss incurred when θ_i is the true state of nature, and $\hat{\theta}_i$ is the point estimate[4] of θ_i.

The minimum expected loss (MEL) rule[5] is to choose $\hat{\theta}_i$ such that

$$\int_{\Omega_i} L(\theta_i,\hat{\theta}_i)p(\theta_i|y,M_i)d\theta_i$$

is minimized. If this integral is finite, then the Fubini Theorem may be applied, and it can be shown that $\hat{\theta}_i$ is a MEL estimate iff it is a Bayes estimate of θ_i. As such,

---

4. The selection of $\hat{\theta}_i$ amounts to taking some action, a.
5. See Zellner (1971), pp.24-26; Geisel, op.cit., pp.9-11.

$\hat{\theta}_i$ minimizes average risk. However, in some cases the M.E.L. estimate exists when the Bayes estimate does not, and so in this thesis we concentrate on M.E.L. decision rules.[6] As it is well known[7], if L is positive definite quadratic then the M.E.L. estimate, $\hat{\theta}_i$, is the mean of the conditional posterior p.d.f. for θ_i:

$$\hat{\theta}_i = \mathcal{E}(\theta_i | y, M_i).$$

Now, let μ'_r denote the rth moment about zero for θ_i. So,

$$\mu'_r = \int_{\Omega_i} \theta_i^r p(\theta_i | y) d\theta_i$$

$$= \int_{\Omega_i} \theta_i^r [\Sigma p(\theta_i | y, M_i) p(M_i | y)] d\theta_i$$

$$= \sum_i p(M_i | y) \int_{\Omega_i} \theta_i^r p(\theta_i | y, M_i) d\theta_i$$

$$= \sum_i p(M_i | y) \mu'_{ri} , \quad \text{say.} \qquad \text{II.III.1}$$

Thus, under the above loss function the marginal M.E.L. estimate of θ_i is

$$\theta_i^* = \mathcal{E}(\theta_i | y) = \sum_i p(M_i | y) \mathcal{E}(\theta_i | y, M_i) \qquad \text{II.III.2}$$

Let μ_r denote the rth moment about the mean for θ_i. Then it is well known that

$$\mu_2 = \mu'_2 - (\mu'_1)^2 \qquad \text{II.III.3}$$

---

6. Here we are following Thornber, op.cit.; Geisel, op.cit.; and Gaver, op.cit..
7. For example, see Zellner (1971), p. 24.

Combining II.III.1 and II.III.3,

$$\mu_2 = \sum_i p(M_i|y)[\mu_{2i} + (\mu'_{1i})^2] - [\sum_i p(M_i|y)\mu'_{1i}]^2 \qquad \text{II.III.4}$$

Thus,

$$V(\theta_i|y) = \sum_i p(M_i|y)[V(\theta_i|y,M_i) + (\mathcal{E}(\theta_i|y,M_i))^2]$$

$$- [\sum_i p(M_i|y)\mathcal{E}(\theta_i|y,M_i)]^2 \qquad \text{II.III.5}$$

IV. PREDICTION

Let $y_F \in Y_F$ be a vector of as yet unobserved (future) observations for the endogenous variable of the stochastic process. Then the <u>predictive p.d.f.</u> for y_F is

$$p(y_F|y) = \sum_i p(y_F|y, M_i)p(M_i|y), \qquad \text{II.IV.1}$$

where $p(M_i|y)$ is obtained in II.II.4,

$$p(y_F|y,M_i) = \int_{\Omega_i} p(y_F|\theta_i,M_i,y)p(\theta_i|y,M_i)d\theta_i \qquad \text{II.IV.2}$$

and $p(\theta_i|y, M_i)$ is from II.II.3.

In this case, the loss function maps $(Y_F \times A)$ to $[0,\infty)$, and $L(y_F, \hat{y}_F)$ denotes the loss incurred when $\hat{y}_F$ is chosen as the point estimate of the unknown y_F. Again, the M.E.L. rule leads to a choice of $\hat{y}_F$ such that

$$\int_{Y_F} L(y_F,\hat{y}_F)p(y_F|y,M_i)dy_F$$

is minimized[8], and results analogous to II.III.2 and II.III.5 apply here.

V. MODEL COMPARISONS

Let $\bar{M}_j$ denote the action of selecting the jth. model from $\mathcal{M}$. In this case L maps $(\mathcal{M} \times A)$ to $[0,\infty)$. Since M is assumed to be countable, the MEL and Bayes rules are equivalent[9], and M_i is chosen from $\mathcal{M}$ iff

$$\sum_r L(M_r,\bar{M}_i)p(M_r|y) < \sum_r L(M_r,\bar{M}_j)p(M_r|y) \qquad \text{II.V.1}$$

for all $j\neq i$.

If attention focuses on just two of the models in $\mathcal{M}$, then the B.P.O. relating M_k and M_j are

$$\{p(M_i|y)/p(M_j|y)\} = \{p(M_k)/p(M_j)\}\{p(y|M_k)/p(y|M_j)\}, \qquad \text{II.V.2}$$

which depends on the prior odds and the ratio of weighted likelihood functions, the latter being obtained from II.II.1. The B.P.O. are independent of $p(y)$, so there is no need to specify the full extent of $\mathcal{M}$ if only M_k and M_j are of interest. Although the B.P.O. are unaffected by the presence or absence of other models as long as the prior odds are unaffected, clearly any individual posterior

8. See Zellner (1971), p.30.
9. See Geisel, op.cit., pp.9-11.

probability depends on the other models, through $p(y)$.

An interesting large-sample approximation can be obtained by re-writing II.II.1 as

$$p(y|M_i) = \int_{\Omega_i} \ell(\theta_i|y,M_i)p(\theta_i|M_i)d\theta_i,$$

and then following Lindley (1961) and applying a Taylor expansion to the likelihood function:

$$p(y|M_i) \simeq \int_{\Omega_i} \ell(\tilde{\theta}_i|y,M_i)p(\theta_i|M_i)d\theta_i, \qquad \text{II.V.3}$$

where $\tilde{\theta}_i$ is the Maximum Likelihood estimate of θ_i, and only the first term of the expansion is retained. Then, if $p(\theta_i|M_i)$ is a proper[10] prior p.d.f., the large-sample approximation in II.V.3 becomes:

$$p(y|M_i) \simeq \ell(\tilde{\theta}_i|y,M_i), \qquad \text{II.V.4}$$

so that:

$$\{p(M_k|y)/p(M_j|y)\} \simeq \{p(M_k)/p(M_j)\} \cdot \{\ell(\hat{\theta}_k|y,M_k)/\ell(\hat{\theta}_j|y,M_j)\} \qquad \text{II.V.5}$$

Thus, the prior odds transform the usual likelihood ratio into an approximate posterior odds ratio, regardless of the form of the (proper) prior p.d.f.'s for the

10. That is, if $\int_{\Omega_i} p(\theta_i|M_i)d\theta_i = 1$.

parameters in each model.

In the two-model case, M_k is preferred to M_j by the MEL rule iff

$$\{p(M_k|y)/p(M_j|y)\} > \{L(M_j,\overline{M}_k)/L(M_k,\overline{M}_j)\} \qquad \text{II.V.6}$$

where the L.H.S. of II.V.6 is computed from II.II.4, or may be <u>approximated</u> (in large samples) by II.V.5.

Thus, if the loss function is symmetric, so that

$$L(M_j,\overline{M}_k) = L(M_k,\overline{M}_j) \quad ; \quad \text{for all } j,k;$$

then the MEL rule leads to choosing the model with the highest posterior probability.

Finally, results established by Lempers and by Geisel[12] indicate that $\operatorname{plim}_{n\to\infty} \{p(M_i|y)\} = 1$, and $\operatorname{plim}_{n\to\infty} \{p(M_j|y)\} = 0$, for all $j \neq i$, when M_i is the true model and n is the number of sample observations.

These results of Bayesian inference summarized here form the basis for the analysis in the remainder of the thesis. The presentation here is brief in view of the detailed discussions given in the various references cited above.

11. See Zellner and Palm (1973), pp.23-24.
12. See Lempers, <u>op.cit.</u>, pp.37-41; and Geisel, <u>op.cit.</u>, pp.22-23.

CHAPTER III

A M.S.E. COMPARISON OF TWO M.E.L. ESTIMATORS FOR THE MULTIPLE REGRESSION MODEL

I. INTRODUCTION

In this Chapter we investigate some aspects of the sampling properties of two M.E.L. estimators for the coefficient vector, β, in the usual multiple regression model. These estimators are $\hat{\beta}$, the Ordinary Least Squares (O.L.S.) estimator; and $\tilde{\beta}$, the Bayes estimator based on a Natural-Conjugate prior p.d.f. and a quadratic loss function (N.C.B.).

Thus, some may view the contents of this Chapter as being only semi-Bayesian, since the strict Bayesian view places relatively little emphasis on the sampling properties of point estimators.[1] However, some analysts are concerned with such properties, as is clear from the recent contributions by Zellner and Vandaele (1972), and Smith (1973). Although M.E.L. and Bayes estimators are usually obtained for reasons unrelated to their sampling properties, we believe that these properties are interesting, none the less.

The basis of comparison used here is[2] generalized (matrix) M.S.E.. Thus, $\tilde{\beta}$ is said to be "preferred" to $\hat{\beta}$ here if [M.S.E.($\hat{\beta}$) - M.S.E.($\tilde{\beta}$)] is a positive semi-definite (p.s.d.) matrix. This is equivalent to requiring that M.S.E.($\eta'\tilde{\beta}$) $\leq$ M.S.E.($\eta'\hat{\beta}$), for all non-zero k-vectors, η. Of course,

---

1. For example, see Tiao and Box (1973).
2. See Goldberger (1964), p.129.

other comparative bases are possible - Wallace(1972) uses a weaker M.S.E. criterion which is a special case of that just noted, and Thornber(1967) discusses the merits of absolute and average risk criteria. B.P.O. could also be used here.

The choice between $\tilde{\beta}$ and $\hat{\beta}$ may be viewed as a choice between amounts of prior information to be used when estimating β, since $\hat{\beta}$ may be interpreted as a M.E.L. estimator based on a diffuse[3] prior p.d.f. and a quadratic loss function. Although $\hat{\beta}$ is inadmissible under quadratic loss in models of three or more regressors(see Stein(1960) and Sclove(1968), for example), it is frequently adopted in practice. Thus it is interesting to compare its sampling properties with those of $\tilde{\beta}$.

The two estimators are equivalent in large samples, since under very general regularity conditions $\tilde{\beta}$ converges to the Maximum Likelihood (M.L.) estimator[4]; here $\hat{\beta}$. Thus, under these conditions both $\hat{\beta}$ and $\tilde{\beta}$ are consistent. However, in finite samples their properties generally differ, as is made apparent in this Chapter. The situation is complicated by the fact that $\tilde{\beta}$ is both non-linear and biased, while $\hat{\beta}$ is the minimum variance linear unbiased estimator of β.

If one of $\hat{\beta}$ or $\tilde{\beta}$ is to be chosen[5] on the basis of matrix M.S.E. then a practical test procedure may be required. This possibility is considered in Section VI, and there we abstract from the well-known problem of pre-testing bias.[6]

---

3. For a discussion of diffuse prior p.d.f.'s see Zellner (1971), pp.41-53. For a discussion of the relationship between M.E.L. and Bayes estimators see Section III of Chapter II.
4. See Zellner, <u>op. cit.</u>, pp. 31-34.
5. We do not consider the possibility of "mixed strategies": <u>i.e.</u> the possibility of using $\bar{\beta}=\omega\beta+(1-\omega)\beta$; $0<\omega<1$.
6. See Bancroft(1944), and Judge <u>et al</u>.(1974).

Finally, we focus attention only on point estimators, but it must be emphasised that an important feature of M.E.L. (and Bayes) procedures is that they lead to a complete posterior p.d.f., containing all relevant information, on which interval estimates, predictions, etc. may be based.

II. PRELIMINARY RESULTS

Consider the usual linear regression model:

$$y = X\beta + u \qquad \text{III.II.1}$$

where:

(i) y is an n-vector of observations on the dependent variable.

(ii) X is $(n \times k)$; a matrix of n observations on k fixed explanatory variables.

(iii) β is a k-vector of unknown parameters.

(iv) u is an n-vector of unknown disturbances, with u_t $NID(0,\sigma^2)$; $t = 1,2,\ldots,n$.

The O.L.S. estimator of β is

$$\hat{\beta} = (X'X)^{-1}X'y \qquad \text{III.II.2}$$

with[7]

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} \qquad \text{III.II.3}$$

The likelihood function for III.II.1 is

---

7. Throughout this Chapter, $V(\cdot)$ denotes a covariance matrix; var. $(\cdot)$ denotes a scalar variance; and $\mathcal{E}(\cdot)$ denotes expectation, all with respect to the space Y, given the model and the parameters.

$$\ell(\beta,\sigma|y) \propto \sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-X\beta)'(y-X\beta)\} \qquad \text{III.II.4}$$

We consider a Bayesian analysis of III.II.1, first under prior ignorance, and secondly when some prior information is available. In the former case, Jeffreys' diffuse prior p.d.f. is adopted:

$$p(\beta,\sigma) = p(\beta).p(\sigma)$$

where

$$p(\beta).d\beta \propto d\beta \; ; \quad -\infty < \beta_i < \infty; \; i = 1,2,\ldots,k \qquad \text{III.II.5}$$

$$p(\sigma).d\sigma \propto d\sigma/\sigma \; ; \quad 0 < \sigma < \infty. \qquad \text{III.II.6}$$

Although the p.d.f. in III.II.6 is improper, an appeal to the probability axioms of Rényi (1970) permits its use in Bayes' Theorem:

$$p(\beta,\sigma|y) \propto \sigma^{-(n+1)}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-X\beta)'(y-X\beta)\} \qquad \text{III.II.7}$$

and,

$$p(\beta|y) \propto \{\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\}^{-n/2} \qquad \text{III.II.8}$$

where

$$\nu = (n-k), \text{ and } \nu s^2 = (y-X\hat{\beta})'(y-X\hat{\beta}).$$

Under a positive definite (p.d.) quadratic loss function, $\hat{\beta}$ is the M.E.L. estimator of β in this case.

One way of introducing prior information about the parameters of III.II.1 is through the Natural-Conjugate prior p.d.f., here a Normal-inverted Gamma density[8]:

8. The notation here follows Zellner, <u>op. cit.</u>, pp.75-76.

$$p(\beta,\sigma) = p(\beta|\sigma).p(\sigma)$$

where

$$p(\beta|\sigma) \propto |A|^{\frac{1}{2}}\sigma^{-k}\exp\{-\tfrac{1}{2}\sigma^{-2}(\bar{\beta}-\beta)'A(\bar{\beta}-\beta)\} \qquad \text{III.II.9}$$

$$p(\sigma) \propto \sigma^{-(w+1)}\exp\{-\tfrac{1}{2}\sigma^{-2}wc^{2}\} \; ; \qquad w > 0 \qquad \text{III.II.10}$$

By Bayes' Theorem,

$$p(\beta,\sigma|y) \propto \sigma^{-(m+k+1)}$$

$$.\exp\{-\tfrac{1}{2}\sigma^{-2}[mq^{2}+(\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta})]\} \qquad \text{III.II.11}$$

and

$$p(\beta|y) \propto \{mq^{2} + (\beta-\tilde{\beta})'(A+X'X)(\beta-\tilde{\beta})\}^{-(m+k)/2} \qquad \text{III.II.12}$$

where:

$$m = n+w \; ; \qquad \tilde{\beta} = (A+X'X)^{-1}(A\bar{\beta}+X'y);$$

$$mq^{2} = \{wc^{2} + y'y + \bar{\beta}'A\bar{\beta} - \tilde{\beta}'(A+X'X)\tilde{\beta}\} \qquad \text{III.II.13}$$

and A must be positive definite symmetric (p.d.s) for III.II.9 to be a <u>proper</u> p.d.f.,

In this case under a p.d. quadratic loss function $\tilde{\beta}$ in III.II.13 is the M.E.L. estimator of β.

The estimator $\tilde{\beta}$ is biased (and non-linear) in general, while $\hat{\beta}$ is unbiased (and linear). However, we shall show that any linear combination of the elements of the latter has larger sampling variance than has the corresponding linear combination of the elements of the former. Thus, there will be some regions of the parameter space in which any linear combination of the elements of $\tilde{\beta}$ has smaller M.S.E. than has the corresponding linear combination of the elements of $\hat{\beta}$.

For the purposes of this Chapter (and for part of the analysis in Chapter IV) we shall describe $\tilde{\beta}$ as being "preferred" to $\hat{\beta}$ in terms of its sampling properties[9] if the above M.S.E. relationship is satisfied. This strong M.S.E. criterion is used in a similar way in other comparative studies, such as those of Toro-Vizcarrondo and Wallace (1968), and Griffiths (1973).

III. COMPARING $\hat{\beta}$ AND $\tilde{\beta}$

Clearly, $\tilde{\beta}$ differs from $\hat{\beta}$ by the prior information reflected[10] in A and $\bar{\beta}$. This information decreases as $A \to 0$ and $w \to 0$, for then $p(\beta,\sigma|y)$ in III.II.11 differs in the limit from $p(\beta,\sigma|y)$ in III.II.8 only[11] in the power of σ. Further, $\tilde{\beta}$ converges to $\hat{\beta}$ in large samples, for then the sample information dominates that in the prior p.d.f., unless the latter is totally "dogmatic", <u>i.e.</u> $A^{-1} = 0$. (In this case $\tilde{\beta} = \bar{\beta}$ and $p(\beta|\sigma)$ in III.II.9 is an improper density.) It is also interesting to note that $\tilde{\beta}$ may be expressed as

$$\tilde{\beta} = \bar{\beta} + [I - (A+X'X)^{-1}A](\hat{\beta}-\bar{\beta}),$$

which is of Stein-like form. Clearly, if $\bar{\beta} = \hat{\beta}$, then $\tilde{\beta} = \hat{\beta}$.

---

9. Of course, a strict Bayesian would always use $\tilde{\beta}$ in preference to $\hat{\beta}$, since the former incorporates <u>all</u> of the available relevant <u>a priori</u> information.
10. Zellner (1972) shows that the construction of A should depend on c and w.
11. See Lempers (1971), Chapter 2.

If $\beta^* = (\bar{\beta}-\beta)$,

and $W = (A+X'X)^{-1}$,

then

$$\text{Bias } (\tilde{\beta}) = WA\beta^* \qquad \text{III.III.1}$$

so that Bias $(\tilde{\beta}_i)$ depends not only on β_i^*, but also on β_j^* for all $j \neq i$. Further,

$$V(\tilde{\beta}) = \sigma^2 WX'XW' \qquad \text{III.III.2}$$

Thus, $\tilde{\beta}$ is unbiased iff $\bar{\beta} = \beta$. However, as is well known, $\hat{\beta}$ is unbiased with covariance matrix as in III.II.3.

Using the matrix M.S.E. concept noted earlier, a strong criterion [12] for an arbitrary estimator, b_1, to be preferred to another arbitrary estimator, b_2, is:

$$\text{M.S.E.}(\eta' b_1) \leq \text{M.S.E.}(\eta' b_2), \qquad \text{III.III.3}$$

for all $\eta \neq 0$. This is equivalent to

$$\mathcal{E}(b_2-\beta)(b_2-\beta)' = \mathcal{E}(b_1-\beta)(b_1-\beta)' + D$$

for a $(k \times k)$ p.s.d. D; or:

$$V(b_2) + \text{Bias }(b_2).\text{Bias}(b_2)'$$

$$= V(b_1) + \text{Bias}(b_1).\text{Bias}(b_1)' + D \qquad \text{III.III.4}$$

12. Wallace, op.cit., suggests a weaker criterion, replacing III.III.3 by: $\sum_i \text{M.S.E.}(b_{1i}) \leq \sum_i \text{M.S.E.}(b_{2i})$. The weak criterion is a special case of the strong one, as is seen by letting $\eta' = (0,\ldots,1,\ldots,0)$ in III.III.3, and summing over all i. The weak criterion is also used by Judge et al. (1973), (1974), and by Bock et al. (1973).

We now show that any linear combination of the elements of $\tilde{\beta}$ always has smaller variance than has the corresponding linear combination of the elements of $\hat{\beta}$.

Proposition III.III.1:

$\Delta = V(\hat{\beta}) - V(\tilde{\beta})$ is a p.d.s. matrix.

Proof: Let η be any non-zero k-vector. Then, from III.II.3 and III.III.2,

$$\eta'\Delta\eta = \eta'\{\sigma^2(X'X)^{-1} - \sigma^2 WX'XW'\}\eta.$$

$$\text{Let } \xi = W'\eta,$$

then:

$$\xi'(W^{-1}\Delta W^{-1})\xi = \sigma^2\xi'\{W^{-1}(X'X)^{-1}W^{-1} - X'X\}\xi,$$

since W is p.d.s. (by[13] Theorems A.2 and A.3).

However,

$$W^{-1}(X'X)^{-1}W^{-1} = A(X'X)^{-1}A + X'X + 2A \qquad \text{III.III.5}$$

since A is symmetric by construction.

Thus,

$$\xi'(W^{-1}\Delta W^{-1})\xi = \sigma^2\xi'\{A(X'X)^{-1}A + 2A\}\xi.$$

Now, A is p.d.s., so it is non-singular, by Theorem A.4, and $A' = A$. Thus, by Theorem A.1, $A(X'X)^{-1}A$ is p.d.s.,

13. Theorems prefixed by the letter "A" appear in Appendix I.

since rank $(X'X) = k$. Further, $\{A(X'X)^{-1}A + 2A\}$ is p.d.s., by Theorem A.3. Since $\sigma^2 > 0$, $(W^{-1}\Delta W^{-1})$ is p.d.s., so Δ is p.d.s. by Theorem A.1. <u>Q.E.D.</u>

Thus, $\hat{\beta}$ has a covariance matrix which "dominates" that of $\tilde{\beta}$, and under certain conditions this may mean that $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ in terms of M.S.E., even though $\tilde{\beta}$ is biased while $\hat{\beta}$ is not. Necessary and sufficient conditions emerge in Proposition III.III.2.

Finally, since Δ is p.d., $\eta'\Delta\eta > 0$ for all $\eta \neq 0$. Equivalently, var.$(\eta'\hat{\beta}) >$ var.$(\eta'\tilde{\beta})$ for all $\eta \neq 0$. In particular, letting $\eta = (0,\ldots,0,1,0,\ldots,0)$, var.$(\hat{\beta}_i) >$ var.$(\tilde{\beta}_i)$, for all i. This result is useful in Part (3) of this Section.

<u>Proposition III.III.2</u>:

A necessary and sufficient condition for $\tilde{\beta}$ to be "preferred" to $\hat{\beta}$ in terms of the strong M.S.E. criterion is that

$$\beta^{*'}\{\sigma^2(X'X)^{-1} + 2\sigma^2A^{-1}\}^{-1}\beta^* \leq 1.$$

<u>Proof</u>: Applying criterion III.III.4 to III.II.3, III.III.1 and III.III.2, $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ in terms of the strong M.S.E. condition iff

$$\sigma^2(X'X)^{-1} = \sigma^2WX'XW' + WA\beta^*\beta^{*'}A'W' + D_1$$

for some p.s.d. $(k \times k)$ matrix D_1. That is, iff

$$\eta'\{\sigma^2(X'X)^{-1} - \sigma^2 WX'XW' - WA\beta^*\beta^{*'}A'W'\}\eta \geq 0,$$

III.III.6

for <u>all</u> non-zero k-vectors, η.

Let $\xi = W'\eta$, then III.III.6 holds iff

$$\xi'\{W^{-1}(X'X)^{-1}W^{-1} - (X'X) - \sigma^{-2}A\beta^*\beta^{*'}A'\}\xi \geq 0,$$

since W is p.d.s.

But, from III.III.5, $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ on the strong M.S.E. criterion iff

$$\lambda^* = \{\sigma^{-2}\xi'A\beta^*\beta^{*'}A'\xi\}/\{\xi'[A(X'X)^{-1}A + 2A]\xi\}$$

$$\leq 1.$$

III.III.7

Finally, III.III.7 holds for <u>all</u> non-zero ξ as defined above iff

$$\lambda = \sup_{(\xi)}.(\lambda^*) \leq 1;$$

that is, iff[14]

$$\lambda = \beta^{*'}\{\sigma^2(X'X)^{-1} + 2\sigma^2A^{-1}\}^{-1}\beta^*$$

$$\leq 1.$$

III.III.8

<u>Q.E.D.</u>

As expected, condition III.III.8 involves the unknown β and σ^2 as well as the known X, A and $\bar{\beta}$. Further, both $(X'X)$ and A are symmetric, so λ is a quadratic form in the

14. See Rao (1965), p.48.

vector $\beta^* = (\tilde{\beta}-\beta)$. The terms $\sigma^2(X'X)^{-1}$ and σ^2A^{-1} in III.III.8 are $V(\hat{\beta})$ and the conditional prior covariance matrix for β, from III.II.9. Also, note a very important feature of the criterion developed in Proposition III.III.2. The result is asymmetric in that it does not follow that $\hat{\beta}$ is preferred[15] to $\tilde{\beta}$ if $\lambda \geq 1$. If $\lambda > 1$ then there are some η for which $\tilde{\beta}$ is preferred to $\hat{\beta}$, and some for which the reverse is true. Thus, there is not a 1 - 1 correspondence between the two conditions:

(i) $\lambda = 1$

(ii) $M.S.E.(\eta'\tilde{\beta}) = M.S.E.(\eta'\hat{\beta})$; for all $\eta \neq 0$..

This last fact should be borne in mind when interpreting some of the results in the next Section.

Finally, we consider a property of that value of ξ for which the supremum in Proposition III.III.2 is attained.

Proposition III.III.3:

Let ξ_0 be the value of ξ for which λ is attained. Then, if $\eta_0 = W^{-1}\xi_0$, $\text{Bias}(\eta_0'\tilde{\beta}) = \lambda > 0$.

Proof: As a corollary to the theorem cited in footnote 14,

$$\xi_0 = [\sigma^2(X'X)^{-1}A + 2\sigma^2 I]^{-1}\beta^*,$$

so,

$$\eta_0 = (A + X'X)[\sigma^2(X'X)^{-1}A + 2\sigma^2 I]^{-1}\beta^*.$$

Thus,

---

15. The argument used in Proposition III.III.2 cannot be used to consider the conditions under which $\hat{\beta}$ is preferred to β on the strong M.S.E. criterion. In that case the step from III.III.7 to III.III.8 cannot be made. The result of Proposition III.III.2 is asymmetric.

$$\begin{aligned}\text{Bias}(\eta_0'\tilde{\beta}) &= \eta_0' WA\beta^* \\ &= \beta^{*'}[\sigma^2(X'X)^{-1} + 2\sigma^2 A^{-1}]^{-1}\beta^* \\ &= \lambda \\ &> 0\ ; \quad \text{for all } \beta^* \neq 0.\end{aligned}$$

<u>Q.E.D.</u>

The fact that $\text{Bias}^2(\eta_0'\tilde{\beta}) = \lambda^2$ is exploited in Part (2) of the next Section. Further, λ is attained for a <u>particular</u> ξ_0 (or η_0), and for this particular vector the squared bias of the corresponding linear combination of the elements of $\tilde{\beta}$ is non-zero. The importance of this is clarified at the start of the next Section.

IV. FURTHER ANALYTIC RESULTS

The fact that λ in III.III.8 depends on unknown β and σ^2 poses some difficulties for the use of this criterion in applications, if some practical test procedure is desired. Toro-Vizcarrondo and Wallace show that in their problem $\frac{1}{2}\lambda$ is the non-centrality parameter of a non-central F-distribution, so that a formal test is feasible. In the absence of such a test in his problem, Griffiths[16] chooses to explore part of the parameter space numerically for some simple special cases.

We consider the possibility of a test statistic in Section VI, but before turning to this we present some analytic results relating to the general multiple regression model, III.II.1 - in particular we investigate the influences of the parameters of the prior distribution and of the data

16. See Griffiths, <u>op. cit</u>., pp.8-14.

sample. Of primary interest are the influences on λ itself, but attention is also paid to the effects on the sampling properties of $\tilde{\beta}$ and $\hat{\beta}$, where applicable.

Throughout this Section we often consider linear combinations of $\tilde{\beta}$ or $\hat{\beta}$: $\eta'\tilde{\beta}$, $\eta'\hat{\beta}$; for non-zero k-vectors, η. This scalarizes the analysis so that the results may be shown diagrammatically. Clearly, there is nothing to be gained by considering those η for which $\mathrm{Bias}(\eta'\tilde{\beta}) = 0$, for in that case the result of Proposition III.III.1 implies that $\tilde{\beta}$ is always preferred to $\hat{\beta}$ on the basis of M.S.E.. Thus, the analysis proceeds for those η such that $\eta'WA\beta^* \neq 0$. This restriction is not at all severe, since for particular A, X, β and $\bar{\beta}$, the class of η for which $\eta'WA\beta^* = 0$ is a set of measure zero. Finally, the result of Proposition III.III.3 indicates that this degenerate class of η may be totally disregarded when investigating the properties of λ, for the latter is defined for a value of η outside of this class.

(1) <u>Specification of $\bar{\beta}$</u>

The conditional prior mean vector, $\bar{\beta}$, affects M.S.E.($\tilde{\beta}$) and λ only by affecting Bias($\tilde{\beta}$). Clearly, from III.III.2, $V(\tilde{\beta})$ is independent of $\bar{\beta}$. Let θ be an arbitrary positive scalar parameter that scales all of the elements of $\beta^* = (\bar{\beta}-\beta)$ equally:

$$\beta^* = \theta b^*,$$

for $(k \times 1)$ non-zero b^*.

Let

$$B = \mathrm{Bias}(\tilde{\beta}).\mathrm{Bias}(\tilde{\beta})'$$

$$= \theta^2 WAb^*b^{*'}A'W' \qquad \text{III.IV.1}$$

Now, $(\beta^* \beta^{*\prime})$ is p.s.d. by Theorem A.9, so $(b^* b^{*\prime})$ is also p.s.d., since $\theta > 0$. Applying Theorem A.1 twice, B is a p.s.d. matrix. Now, from III.IV.1,

$$\underset{\theta \to 0}{\text{Limit}}(B) = 0, \qquad \underset{\theta \to 0}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = V(\tilde{\beta}),$$

and

$$\underset{\theta \to \infty}{\text{Limit}}(B) = \underset{\theta \to \infty}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = \infty.$$

For the purposes of depicting these results diagrammatically it is helpful to consider the scalar quantity

$$\begin{aligned} \bar{B} &= \text{Bias}^2(\eta' \tilde{\beta}) \\ &= \theta^2 (\eta' W A b^*)^2 \end{aligned} \qquad \text{III.IV.2}$$

Limits similar to those above also apply to $\bar{B}$. Further, $\bar{B}$ is a simple quadratic in θ provided that $(\eta' W A b^*) \neq 0$; i.e. $(\eta' W A \beta^*) \neq 0$. For one such η and for particular X, A, β, $\bar{\beta}$ and σ^2, the above results are depicted in Figure III.IV.1.

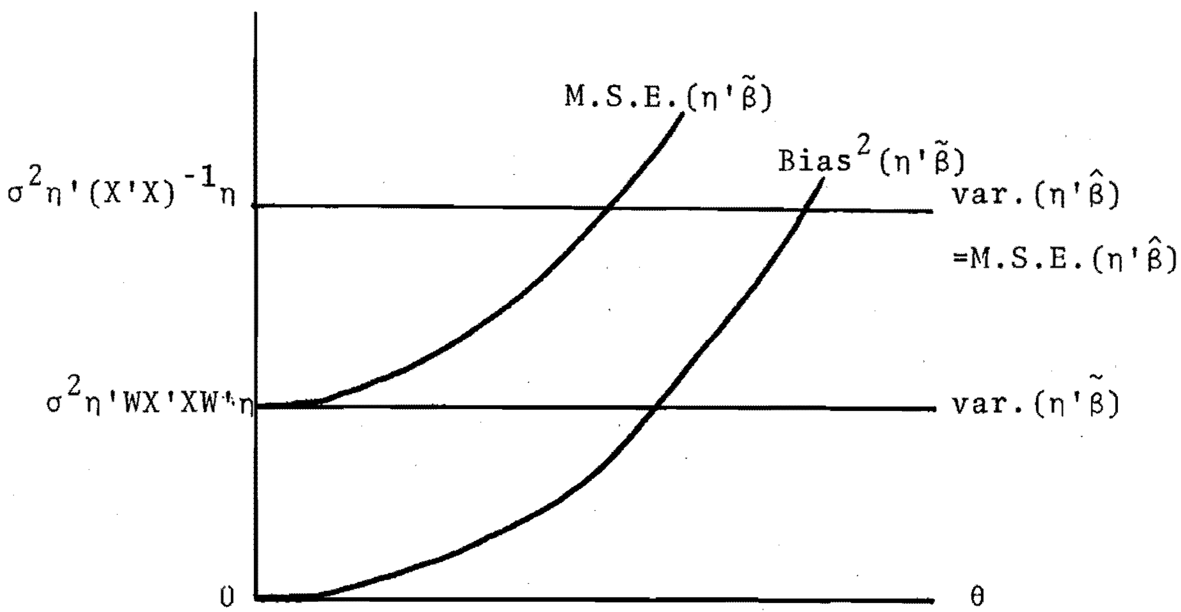


Figure III.IV.1: Effects of varying θ.

Thus, <u>ceteris paribus</u>, by choosing $\bar{\beta}$ sufficiently close to β it is possible for $\tilde{\beta}$ to be "preferred" to $\hat{\beta}$ in the strong sense defined earlier in the Chapter. Of course, β itself is always unknown. A similar conclusion is reached by Smith, using a weak M.S.E. criterion in the context of a Bayesian analysis of an hierarchical linear model.

This particular feature of the way in which the choice of $\bar{\beta}$ affects the sampling properties of $\tilde{\beta}$ in relation to those of $\hat{\beta}$ is highlighted by considering the relationship between λ and θ.

First, since λ in III.III.8 is a quadratic form in β^*, if $\bar{\beta} = \beta$ then $\lambda = 0$ and $\tilde{\beta}$ is "preferred" to $\hat{\beta}$, as expected. Clearly, from III.III.1, Bias$(\tilde{\beta}) = 0$ if $\bar{\beta} = \beta$, so as $\theta \to 0$, the comparison between $\tilde{\beta}$ and $\hat{\beta}$ amounts to a comparison of their covariance matrices, and Proposition III.III.1 takes effect.

Secondly,

$$(\partial\lambda/\partial\beta^*) \quad = \quad 2\{\sigma^2(X'X)^{-1} + 2\sigma^2A^{-1}\}^{-1}\beta^* \qquad \text{III.IV.3}$$

so λ is continuous in β^*. Thus, there exists a continuum of β^* values (or values of θ) such that $\lambda \leq 1$ is satisfied. Thus, <u>ceteris paribus</u>, choosing $\bar{\beta}$ so that $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ may be accomplished in an infinity of ways. Again, this is reflected in Figure III.IV.1.

Further, since III.III.8 may be written

$$\lambda \quad = \quad \theta^2 b^{*\prime}\{\sigma^2(X'X)^{-1} + 2\sigma^2A^{-1}\}^{-1}b^*,$$

λ is a simple quadratic in θ, for all p.d.s. A and $(X'X)$, and $b^* \neq 0$. (Of course, $\lambda = 0$ for all A and X if $b^* = 0$.)

Thus, there exists a unique θ_0 such that $\lambda = 1$:

$$\theta_0 = \{b^{*'}[\sigma^2(X'X)^{-1} + 2\sigma^2A^{-1}]^{-1}b^*\}^{-1/2}, \qquad \text{III.IV.4}$$

which is positive for $b^* \neq 0$, as shown in Figure III.IV.2.

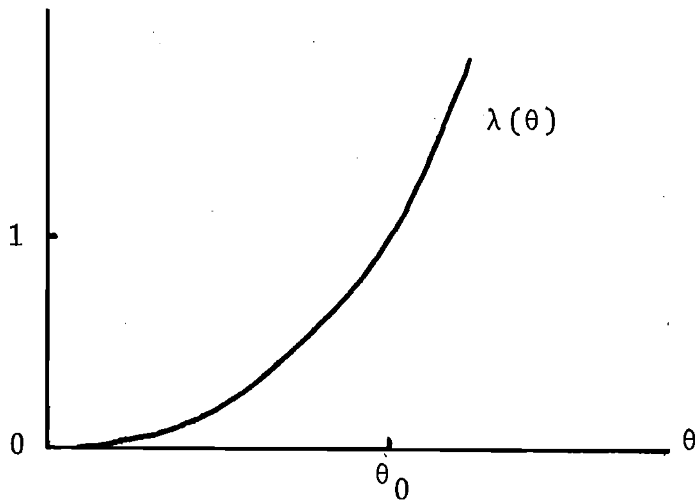


Figure III.IV.2: λ as a function of θ.

The above results concerning λ apply, of course, for the particular vector, ξ_0, for which λ is attained. Although λ is our main concern, the relationship between λ^* and θ is also of some interest. From III.III.7,

$$\lambda^* = \{\theta^2\sigma^{-2}\xi'Ab^*b^{*'}A'\xi\}/\{\xi'[A(X'X)^{-1}A + 2A]\xi\} \qquad \text{III.IV.5}$$

Now, the denominator in III.IV.5 is strictly positive for all ξ, since A and $(X'X)^{-1}$ are p.d.s.. Thus, λ^* is a simple quadratic in θ as long as $(\xi'Ab^*)^2 \neq 0$. For particular A, β and $\bar{\beta}$ there is an infinity of ξ for which $(\xi'Ab^*) \neq 0$. Further, since $\xi = W'\eta$, the required restriction is that $(\eta'WAb^*) \neq 0$, and the relevance of the latter has been discussed already.

Taking account of the proof of Proposition III.III.2, it is not surprising that the restriction required for a strong relationship between λ^* and θ is that required for a strong relationship between $\bar{B}$ and θ.

The results of this sub-section are summarized in Figure III.IV.1, and again it may be emphasized that $\tilde{\beta}$ will be "preferred" to $\hat{\beta}$ if $\bar{\beta}$ is chosen sufficiently close to β, <u>ceteris paribus</u>, for p.d.s. A.

(2) <u>Specification of A</u>

The matrix A affects M.S.E.$(\tilde{\beta})$ and λ through both $V(\tilde{\beta})$ and Bias$(\tilde{\beta})$. Let ϕ be a positive scalar parameter which scales <u>all</u> of the elements of A^{-1} uniformly. Thus, all of the conditional prior covariances and variances of the elements of β are affected equally, and $A^{-1} = \phi A_0{}^{-1}$, where A_0 is also p.d.s.. This approach is also used by Geisel (1970; Ch.3). Clearly, $\underset{\phi \to 0}{\text{Limit}}(\tilde{\beta}) = \bar{\beta}$, and $\underset{\phi \to \infty}{\text{Limit}}(\tilde{\beta}) = \hat{\beta}$. Now, consider the behaviour of M.S.E.$(\tilde{\beta})$ as ϕ varies.

<u>Proposition III.IV.1</u>:

(i) $$\underset{\phi \to 0}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = \beta^* \beta^{*\prime}$$

(ii) $$\underset{\phi \to \infty}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = \text{M.S.E.}(\hat{\beta}) = \sigma^2 (X'X)^{-1}.$$

<u>Proof</u>:

(i) $$\text{Bias}(\tilde{\beta}) = (A_0 + \phi X'X)^{-1} A_0 \beta^*$$

$$V(\tilde{\beta}) = \sigma^2 \phi^2 (A_0 + \phi X'X)^{-1} (X'X)(A_0 + \phi X'X)^{-1}.$$

So,

$$\underset{\phi \to 0}{\text{Limit}}\{\text{Bias}(\tilde{\beta})\} = \beta^*; \quad \underset{\phi \to 0}{\text{Limit}}\{V(\tilde{\beta})\} = 0,$$

and,

$$\underset{\phi \to 0}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = \beta^* \beta^{*'} .$$

(ii) $$\text{Bias}(\tilde{\beta}) = (\phi^{-1}A_0 + X'X)^{-1}\phi^{-1}A_0\beta^*$$

$$V(\tilde{\beta}) = \sigma^2(\phi^{-1}A_0 + X'X)^{-1}(X'X)(\phi^{-1}A_0 + X'X)^{-1}.$$

So,

$$\underset{\phi \to \infty}{\text{Limit}}\{\text{Bias}(\tilde{\beta})\} = 0; \quad \underset{\phi \to \infty}{\text{Limit}}\{V(\tilde{\beta})\} = \sigma^2(X'X)^{-1},$$

and,

$$\underset{\phi \to \infty}{\text{Limit}}\{\text{M.S.E.}(\tilde{\beta})\} = \text{M.S.E.}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Q.E.D.

Corresponding results hold for M.S.E.$(\eta'\tilde{\beta})$.

Now, following the approach adopted in Part (1) of this Section, we consider the specific relationship between ϕ and $\tilde{B}$, and ϕ and var.$(\eta'\tilde{\beta})$. Again, we work with linear combinations of the elements of $\tilde{\beta}$ so that the results may be depicted diagrammatically. These relationships will shed light on the way in which M.S.E.$(\eta'\tilde{\beta})$ is affected by variations in ϕ (or A^{-1}). First, consider the relationship between ϕ and var.$(\eta'\tilde{\beta})$.

Proposition III.IV.2:

Var.$(\eta'\tilde{\beta})$ is monotonic increasing in ϕ, for all non-zero $(k \times 1)$ vectors, η.

Proof: Let $V = \text{var.}(\eta'\tilde{\beta})$

$$= \sigma^2\eta'(\psi A_0 + X'X)^{-1}(X'X)(\psi A_0 + X'X)^{-1}\eta,$$

where:

$$A^{-1} = \phi A_0^{-1} \; ; \; \phi > 0 \; ; \text{ and } \psi = \phi^{-1}.$$

Then,

$$V = \sigma^2 \eta' \{X'X + 2\psi A_0 + \psi^2 A_0 (X'X)^{-1} A_0\}^{-1} \eta.$$

By Theorem A.8,

$$(\partial V/\partial \psi) = -\sigma^2 \eta' \{X'X + 2\psi A_0 + \psi^2 A_0 (X'X)^{-1} A_0\}^{-1}$$
$$.\{2A_0 + 2\psi A_0 (X'X)^{-1} A_0\}$$
$$.\{X'X + 2\psi A_0 + \psi^2 A_0 (X'X)^{-1} A_0\}^{-1} \eta.$$

Repeated application of Theorems A.1, A.2 and A.3 yields:

$$(\partial V/\partial \phi) < 0,$$

since $\psi > 0$ and $\sigma^2 > 0$; and since A_0 and $(X'X)$ are p.d.s.

Finally,

$$(\partial V/\partial \phi) = (\partial V/\partial \psi)(d\psi/d\phi)$$
$$= -\phi^{-2} (\partial V/\partial \psi)$$
$$> 0 \; ; \quad \text{for } \phi > 0.$$

Thus, var.$(\eta'\tilde{\beta})$ is monotonic increasing in ϕ.

Q.E.D.

From the form of $(\partial V/\partial \psi)$ it is clear that $\underset{\phi \to \infty}{\text{Limit}}(\partial V/\partial \phi) = 0$, so from Proposition III.IV.1, var.$(\eta'\tilde{\beta})$ approaches var.$(\eta'\hat{\beta})$ asymptotically from below as $\phi \to \infty$.

However, var.$(\eta'\tilde{\beta})$ is <u>not</u> a convex or concave function of ϕ, as may be ascertained from the expression for $(\partial^2 V/\partial\phi^2)$. The sign of this second derivative varies with ϕ, indicating at least one point of inflexion for finite ϕ. This is supported by the evidence in Section V below, and in Appendix II.

Having established that var.$(\eta'\tilde{\beta})$ increases smoothly with the elements of the conditional prior covariance matrix for β, it might be expected that $\bar{B} = \text{Bias}^2(\eta'\tilde{\beta})$ is monotonic <u>decreasing</u> in ϕ. In fact this need <u>not</u> be the case in general, as is shown in the following Proposition. However, an interesting special case for which $\bar{B}$ is convex in ϕ emerges in Section V.

<u>Proposition III.IV.3</u>:

$\bar{B} = \text{Bias}^2(\eta'\tilde{\beta})$ is <u>not</u> a monotonic function of ϕ, for general A and $k \geq 2$.

<u>Proof</u>: $\bar{B} = (\eta' WA\beta^*)^2$,

so,

$$
\begin{aligned}
(\partial\bar{B}/\partial\phi) &= 2(\eta' WA\beta^*)[(\partial/\partial\phi)(\beta^{*\prime}A'W'\eta)] \\
&= -2\eta'(I+\phi A_0^{-1}X'X)^{-1}(\beta^*\beta^{*\prime}) \\
&\quad .\{(I+\phi X'XA_0^{-1})(A_0(X'X)^{-1}+\phi I)\}^{-1}\eta \\
&= 2\eta' T_1(\beta^*\beta^{*\prime})T_2\eta, \text{ say,}
\end{aligned}
$$

by Theorem A.8.

Although $(\partial\bar{B}/\partial\phi)$ is a quadratic form in an <u>asymmetric</u> matrix, in fact it may be treated as if it were a quadratic form in a <u>symmetric</u>[17] matrix. Now, by Theorem A.6,

---

17. See Theil (1971), p.21.

$$\text{rank}(T_1\beta^*\beta^{*'}T_2) \;=\; \text{rank}(\beta^*\beta^{*'}) \;=\; 1.$$

Thus, for $k \geq 2$, $(T_1\beta^*\beta^{*'}T_2)$ always has less than full column rank, and so it is singular. Thus, by Theorem A.4 it cannot be a p.d. matrix, and so by definition $(\partial\bar{B}/\partial\phi)$ cannot be strictly negative for all non-zero η. Similarly, $(T_1\beta^*\beta^{*'}T_2)$ cannot be negative definite, and so $(\partial\bar{B}/\partial\phi)$ cannot be strictly positive, either. Thus, $\bar{B}$ cannot <u>in general</u> be a monotonic function of ϕ, for $k \geq 2$. <u>Q.E.D.</u>

As is shown in Appendix II, $\bar{B}$ is convex in ϕ when $k = 1$, and the reasons for this will be come apparent from the discussion which now follows, and from that in Section V. Further, for <u>particular</u> η, $\bar{B}$ may be convex in ϕ for $k \geq 2$. An example is for $\eta_0 = W^{-1}\xi_0$, for then $\bar{B} = \lambda^2$, which is convex in ϕ since λ is convex in ϕ, as is shown below in Proposition III.IV.5.

Before attempting to evaluate the behaviour of M.S.E.$(\eta'\tilde{\beta})$ as ϕ varies, we return to the indeterminate result of the last Proposition. This result is more easily appreciated by considering the interpretation problem associated with multicollinearity, as discussed from a Bayesian viewpoint by Leamer (1973).

The source of the indeterminacy is revealed by considering the Curve Décolletage. Our ϕ corresponds to the inverse of Leamer's ρ, so this curve is the locus traced out in k-space by $\tilde{\beta}$ as ϕ varies from zero to infinity. It is also traced by the points of tangency between the elliptical likelihood contours and the elliptical prior contours.

The shape of the Curve Décolletage is determined by the relative positions and shapes of the likelihood and prior contours, and in terms of our notation it reflects the relationship between A and (X'X). Thus, from III.III.1 and III.III.2 this shape is intimately related to both Bias($\tilde{\beta}$) and V($\tilde{\beta}$), and as is emphasised by Leamer it reflects the extent to which multicollinearity poses an interpretation problem.

An extreme collinearity situation appears in Figure III.IV.3 for k=2. In that illustration the situation is highlighted by setting $\hat{\beta} = \beta$, and although this restriction is inessential to the point being made, in fact it is always satisfied "in expectation".

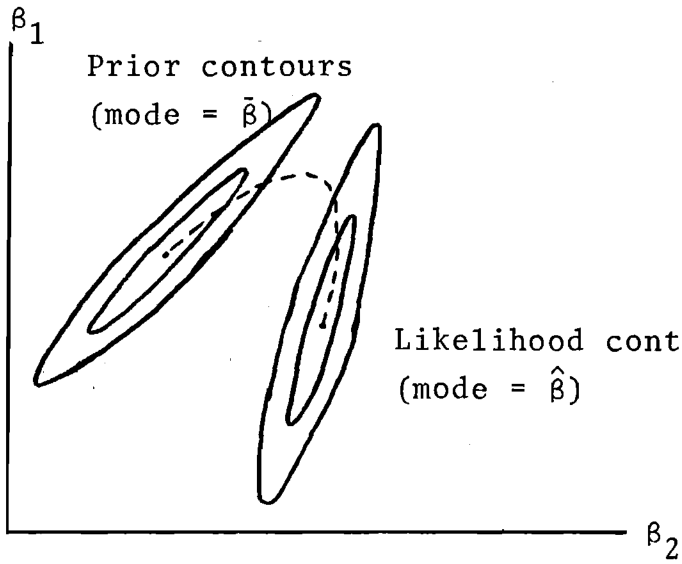


Figure III.IV.3: An extreme curve décolletage.

In Figure III.IV.3, Bias($\tilde{\beta}$) increases and then decreases as ϕ varies from zero to ∞ (i.e. as $\tilde{\beta}$ moves from $\bar{\beta}$ to $\hat{\beta}$). If (X'X) $\propto$ A, then the two sets of elliptical contours are "parallel" and the Curve Décolletage is the line-segment $[\bar{\beta},\hat{\beta}]$. An example of this special case is analysed and illustrated in Section V below.

It is now clear why for general A and arbitrary η, the interaction between A and $(X'X)$ may lead to the indeterminate result in Proposition III.IV.3. In particular, it is interesting that this is closely related to an aspect of the multicollinearity problem.

Further, it is now apparent why, for a particular situation, as ϕ varies over some range M.S.E.$(\eta'\tilde{\beta})$ may possibly increase and then decrease. Such changes are, of course, still governed by the results in Proposition III.IV.1. Three possible situations are summarized for particular (but general) β, $\bar{\beta}$, σ^2 and X in Figures III.IV.4 to III.IV.6.

Of prime interest, of course, are the specific effects of ϕ on λ itself, and we now turn to these.

Proposition III.IV.4:

(i) $\quad \underset{\phi \to 0}{\text{Limit}}(\lambda) = \sigma^{-2}\beta^{*'}(X'X)\beta^* > 0$

(ii) $\quad \underset{\phi \to \infty}{\text{Limit}}(\lambda) = 0.$

Proof:

(i) From III.III.8,

$$\lambda = \beta^{*'}\{\sigma^2(X'X)^{-1} + 2\phi\sigma^2 A_0^{-1}\}^{-1}\beta^*,$$

so

$$\underset{\phi \to 0}{\text{Limit}}(\lambda) = \sigma^{-2}\beta^{*'}(X'X)\beta^*.$$

(ii) Re-arranging III.III.8,

$$\lambda = \phi^{-1}\beta^{*'}\{\sigma^2\phi^{-1}(X'X)^{-1} + 2\sigma^2 A_0^{-1}\}^{-1}\beta^*,$$

so

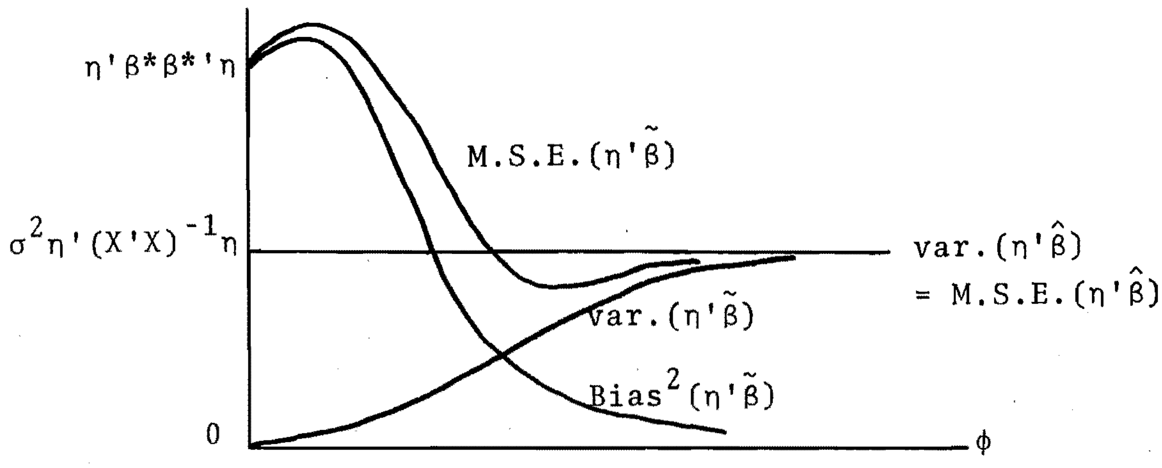


Figure III.IV.4: Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^* > \sigma^2$

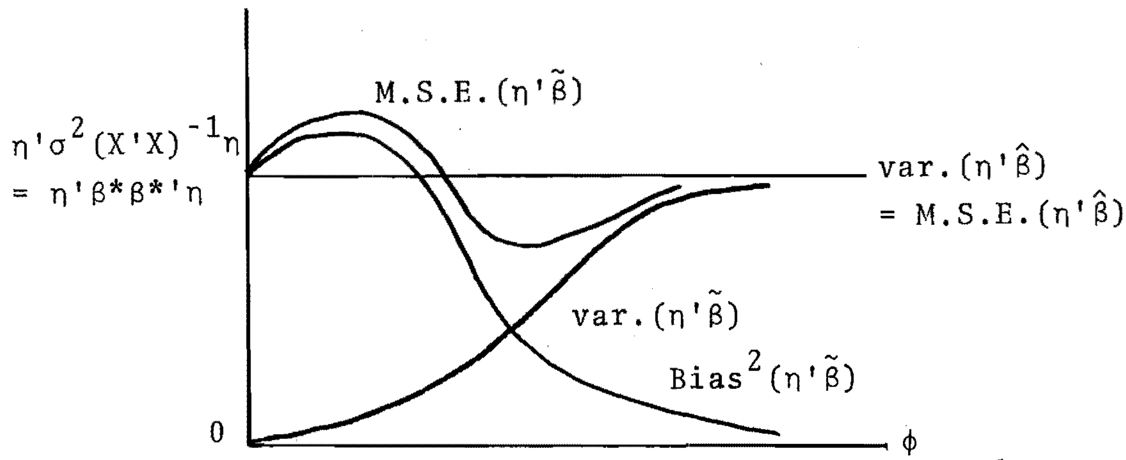


Figure III.IV.5: Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^*=\sigma^2$

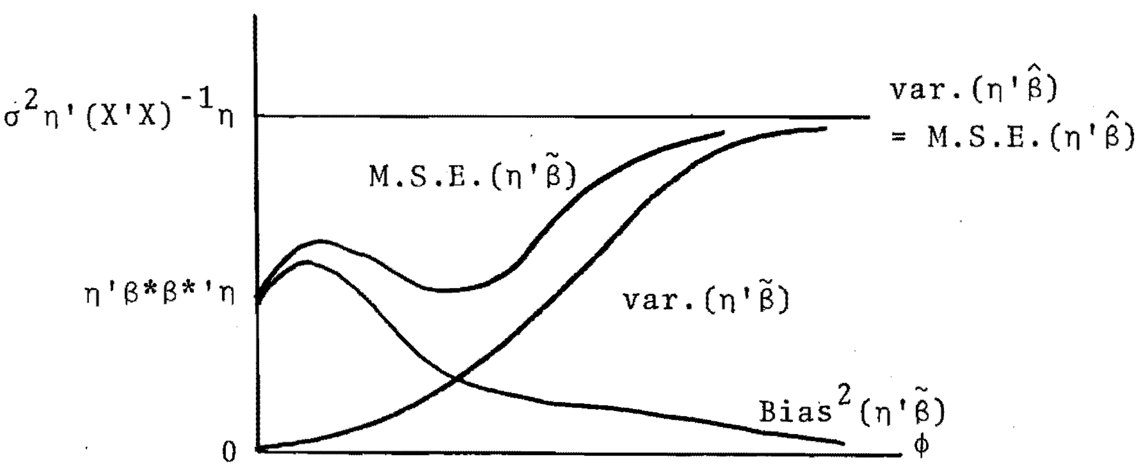


Figure III.IV.6: Effects of varying ϕ when $\beta^{*\prime}(X'X)\beta^*<\sigma^2$

$$\underset{\phi \to \infty}{\text{Limit}}(\lambda) = 0.$$ <u>Q.E.D.</u>

Thus, as the elements of the conditional prior covariance matrix for β increase without limit, $\lambda \to 0$ and $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ in the sense introduced earlier, even though in the limit $\tilde{\beta} = \hat{\beta}$. This point is easily clarified by appealing to Proposition III.IV.1. For arbitrary η, III.III.4 may be re-expressed:

$$\eta'\{V(\hat{\beta}) - V(\tilde{\beta})\}\eta = \eta'\{\text{Bias}(\tilde{\beta}).\text{Bias}(\tilde{\beta})'\}\eta + \eta' D\eta$$

where D is p.s.d.. Thus, $\tilde{\beta}$ is preferred to $\hat{\beta}$ iff

$$\eta'\{\text{Bias}(\tilde{\beta}).\text{Bias}(\tilde{\beta})'\}\eta/\eta'\{V(\hat{\beta}) - V(\tilde{\beta})\}\eta \leq 1 \qquad \text{III.IV.6}$$

Now, by Proposition III.IV.1 the L.H.S. of III.IV.6 is indeterminate as $\phi \to \infty$, but by Proposition III.IV.4, $\underset{\phi \to \infty}{\text{Limit}}(\lambda) = 0$, so in the limit $\tilde{\beta}$ is preferred to $\hat{\beta}$. So, from III.IV.6, the squared bias of $\tilde{\beta}$ approaches zero faster than $V(\tilde{\beta}) \to V(\hat{\beta})$ as $\phi \to \infty$.

Thus, it is natural that as the elements of A^{-1} increase, beyond some point $\tilde{\beta}$ is "preferred" to $\hat{\beta}$, <u>ceteris paribus</u>, and this is consistent with the next Proposition.

<u>Proposition III.IV.5</u>:

λ is a convex function of ϕ.

<u>Proof</u>:

$$\lambda = \beta^{*'}\{\sigma^2(X'X)^{-1} + 2\sigma^2 A^{-1}\}^{-1}\beta^*,$$

so by Theorem A.8,

$$(\partial\lambda/\partial\phi) = -\beta^{*'}\{\sigma^2(X'X)^{-1} + 2\phi\sigma^2 A_0^{-1}\}^{-1}(2\sigma^2 A_0^{-1})$$

$$.\{\sigma^2(X'X)^{-1} + 2\phi\sigma^2 A_0^{-1}\}^{-1}\beta^*$$

$$= -2\sigma^{-2}\beta^{*'}\{(X'X)^{-1}A_0(X'X)^{-1} + 4\phi(X'X)^{-1}$$

$$+ 4\phi^2 A_0^{-1}\}^{-1}\beta^*$$

$$< 0 ; \quad \text{for } \phi > 0,\ \beta^* \neq 0$$

by Theorems A.1, A.2 and A.3.

Further, by Theorem A.8,

$$(\partial^2\lambda/\partial\phi^2) = 2\sigma^{-2}\beta^{*'}\{(X'X)^{-1}A_0(X'X)^{-1} + 4\phi(X'X)^{-1}$$

$$+ 4\phi^2 A_0^{-1}\}^{-1}\{4(X'X)^{-1} + 2\phi A_0^{-1}\}$$

$$.\{(X'X)^{-1}A_0(X'X)^{-1} + 4\phi(X'X)^{-1}$$

$$+ 4\phi^2 A_0^{-1}\}^{-1}\beta^*$$

Repeated application of Theorems A.1, A.2 and A.3 yields:

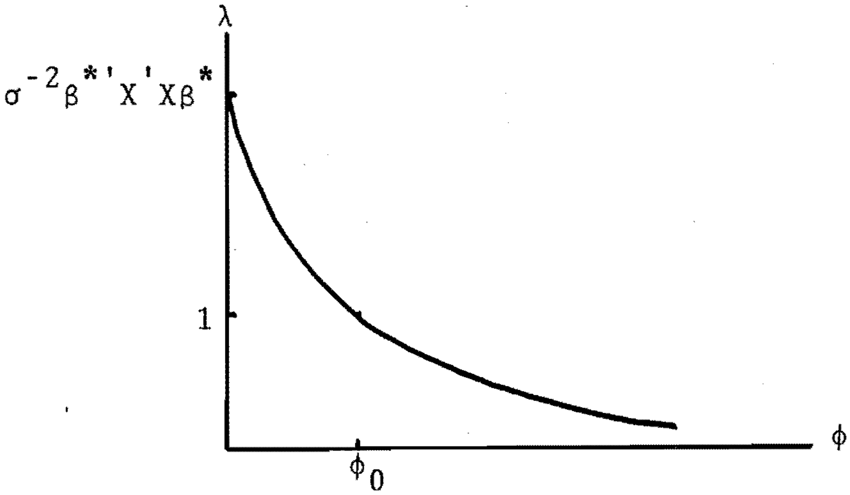
$$(\partial^2\lambda/\partial\phi^2) > 0 ; \qquad \text{for } \phi > 0,\ \beta^* \neq 0$$

since A_0 and $(X'X)$ are p.d.s.. Thus, λ is a convex function of ϕ. Q.E.D.

Thus, if there exists some ϕ_0 such that $\lambda = 1$, then ϕ_0 is unique and $\lambda < 1$ iff $\phi > \phi_0$, $\lambda > 1$ iff $\phi < \phi_0$. As the elements of A^{-1} tend to zero, $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ iff

$$\beta^{*'}(X'X)\beta^{*} \leq \sigma^2 \qquad \text{III.IV.7}$$

So, for particular σ^2, β, $\bar{\beta}$ and X, $\tilde{\beta}$ is "preferred" to $\hat{\beta}$ for <u>all</u> A iff III.IV.7 is satisfied, and in that case ϕ_0 does not exist. The general situation when ϕ_0 does exist is shown in Figure III.IV.7.



<u>Figure III.IV.7</u>: λ as a function of ϕ.

Finally, the fact that λ is convex in ϕ while in general $\text{Bias}^2(\eta'\tilde{\beta})$ is not, (so that M.S.E.$(\eta'\tilde{\beta})$ is not well-behaved with respect to ϕ), is easily explained by reference to Proposition III.III.3. If η_0 is the value of η for which λ is attained, then $\text{Bias}^2(\eta_0'\tilde{\beta}) = \lambda^2$. In this special case $\text{Bias}^2(\eta_0'\tilde{\beta})$ is convex in ϕ since λ^2 is convex in ϕ (as a result of Proposition III.IV.5). Of course, for a general ξ (and hence η), λ^* is not a well-behaved function of ϕ, indeterminacy arising for the same reasons as in Proposition III.IV.3.

(3) Multicollinearity and Precision

Leamer's Bayesian analysis of the interpretation problem associated with multicollinearity has been linked to our own analysis in the previous Part of this Section. Some further related comments appear in the next Section.

Here we consider some aspects of the weak data problem associated with multicollinearity, and compare its effects on $\hat{\beta}$ and $\tilde{\beta}$.

If the exogenous data are perfectly collinear, then the full rank assumption for model III.II.1 is violated and so $\hat{\beta}$ is not defined. (The normal equations have more than one solution, obtainable by appealing to the theory of generalized inverses.) However, A must be p.d.s. for $p(\beta|\sigma)$ in III.II.9 to be proper, so by Theorem A.3 the matrix $(A+X'X)$ is p.d.s., $(A+X'X)^{-1}$ exists by Theorem A.4, and so $\tilde{\beta}$ can be computed as long as A and $\bar{\beta}$ can be specified. In this extreme case although the M.S.E. criterion in III.III.8 cannot be computed, since $(X'X)$ is singular, it is redundant anyway.

A less extreme (and very common) situation arises when the columns of X are collinear to some degree, but not perfectly so. In this case, both $\hat{\beta}$ and $\tilde{\beta}$ are computable, but $\hat{\beta}$ will be imprecise, and so may be $\tilde{\beta}$.

For illustrative purposes, consider the simple regression[18] model:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \; ; \quad t = 1,2,\ldots,n \qquad \text{III.IV.8}$$

18. Here the y and x's are measured as deviations from the sample means.

The O.L.S. estimators are $\hat{\beta}_1$ and $\hat{\beta}_2$, such that:

$$\hat{\beta}_i \sim N(\beta_i, \tau_i \sigma^2) \; ; \qquad i = 1,2$$

where,

$$\tau_i = [(1-r_{12}^2) \sum_t x_{it}^2]^{-1} \; ; \qquad i = 1,2$$

and r_{12} is the simple correlation between x_1 and x_2. The (true and estimated) variances of the above O.L.S. estimators increase without limit as $r_{12} \to 1$. For this simple model, if $A = \begin{pmatrix} V_1 & \alpha \\ \alpha & V_2 \end{pmatrix}$, with $V_1 V_2 > \alpha^2$, then,

$$\underset{r_{12}\to 1}{\text{Limit}} [\text{var.}(\tilde{\beta}_1)] = (\sigma^2 \delta / \gamma^2), \qquad \text{III.IV.9}$$

where:

$$\gamma = V_1 V_2 + V_1 \sum_t x_{1t}^2 + V_2 \sum_t x_{2t}^2 - \alpha^2 - 2\alpha \sum_t x_{1t} x_{2t}$$

$$\delta = V_2^2 \sum_t x_{1t}^2 + \alpha^2 \sum_t x_{2t}^2 - 2\alpha V_2 \sum_t x_{1t} x_{2t}$$

Further,

$$\underset{r_{12}\to 1}{\text{Limit}} [\text{Bias}(\tilde{\beta}_1)] = (\kappa_1 \beta_1^* + \kappa_2 \beta_2^*) / (\kappa_1 + \kappa_2 + \kappa_3) \qquad \text{III.IV.10}$$

where:

$$\kappa_1 = V_1 V_2 + V_1 \sum_t x_{2t}^2 - \alpha^2 - \alpha \sum_t x_{1t} x_{2t}$$

$$\kappa_2 = \alpha \sum_t x_{2t}^2 - V_2 \sum_t x_{1t} x_{2t}$$

$$\kappa_3 = [(V_2 - \alpha) \sum_t x_{1t} x_{2t} + V_2 \sum_t x_{1t}^2 - \alpha \sum_t x_{2t}^2]$$

The limiting expressions for var.$(\tilde{\beta}_2)$ and Bias$(\tilde{\beta}_2)$ are symmetric to those in III.IV.9 and III.IV.10. Thus, as the correlation between x_1 and x_2 increases the variances of the O.L.S. estimates increase without limit, while the variances and biases of the N.C.B. estimates are finite bounded, these upper bounds depending on both the sample and the construction of A and $\bar{\beta}$.

Although it is clear from the discussion following Proposition III.III.1 that var.$(\hat{\beta}_i)$ > var.$(\tilde{\beta}_i)$ for all i and for <u>any</u> degree of collinearity (<u>i.e.</u> for any r_{12} in the simple model), it is now apparent that there is some degree of collinearity above which the var.$(\hat{\beta}_i)$ exceed the corresponding var.$(\tilde{\beta}_i)$ to such an extent that the N.C.B. estimator is "preferred" to $\hat{\beta}$ in terms of the strong M.S.E. criterion. That is, there is some degree of collinearity above which $\lambda \leq 1$ is satisfied, for particular β, A, $\bar{\beta}$, σ^2 and X.

These results relating to the simple model III.IV.8 may be generalized, somewhat informally, to model III.II.1. In particular, note that the inequality III.IV.6 may be rewritten as

$$\beta^{*'}\{\sigma^2(X'X)^{-1}\}^{-1}\beta^* \leq 1, \qquad \text{III.IV.11}$$

and the L.H.S. of III.IV.11 decreases as the elements of $V(\hat{\beta})$ increase. Thus, recalling the discussion in Part (2) of this Section, for particular σ^2, β and $\bar{\beta}$, and for <u>all</u> A, there is some degree of multicollinearity above which $\tilde{\beta}$ is preferred to $\hat{\beta}$ on the basis of M.S.E..

V. A SPECIAL PRIOR P.D.F.

Geisel[19] uses a special form of the Natural-Conjugate prior p.d.f. in III.II.9 and III.II.10, replacing A^{-1} by $h(X'X)^{-1}$, where h is a scalar parameter serving the same purpose as does ϕ in Section IV above.

The limitations of choosing this special form for A are discussed in detail by Geisel, but a relaxation of this strong assumption severely weakens the results that he is able to obtain in his analysis of B.P.O. under decreasing prior information. However, since this prior p.d.f. has received attention in the literature, it is interesting to see how its adoption affects some of the results in the last Section.

Clearly, in this case $A \propto (X'X)$, so the Curve Décolletage is a straight[20] line in k-space. The one-regressor model discussed in Appendix II is a special case of this situation, since there a^{-1} and $\sum_t x_t^2$ are scalars and are therefore always proportional to one another, and so the Curve Decolletage is a line segment on the β-axis.

Thus, in view of the discussion in Part (2) of the last Section, somewhat stronger results with respect to the behaviour of Bias$(\eta'\tilde{\beta})$ and λ^* as h varies might be expected with Geisel's prior p.d.f. than was the case in that Section. These stronger results are merely summarized below.

---

19. See Geisel, <u>op. cit.</u>, p.29 ff.
20. See Leamer, <u>op. cit.</u>, p.374.

First,

$$\tilde{\beta} = (\bar{\beta}+h\hat{\beta})/(1+h) \qquad \text{III.V.1}$$

$$\text{Bias}(\tilde{\beta}) = \beta^*/(1+h) \qquad \text{III.V.2}$$

$$V(\tilde{\beta}) = \sigma^2 h^2 (X'X)^{-1}/(1+h)^2 \qquad \text{III.V.3}$$

$$\lambda^* = \{\sigma^{-2}\xi'X'X\beta^*\beta^{*'}X'X\xi\}/\{(1+2h)\xi'X'X\xi\} \qquad \text{III.V.4}$$

$$\lambda = (\beta^{*'}X'X\beta^*)/\sigma^2(1+2h) \qquad \text{III.V.5}$$

Consider briefly the results corresponding to those in Part (1) of the last Section. Defining θ as before and holding h constant at this stage, we again disregard the (negligible) class of η for which $\text{Bias}(\eta'\tilde{\beta}) = 0$ (i.e. such that $\eta'\beta^* = 0$). Then $\bar{B} \propto \theta^2$ and $\lambda^* \propto \theta^2$. Further, $\lambda \propto \theta^2$ for all η.

Turning to the effects of varying h itself, limits analogous to those of Proposition III.IV.1 hold here. Also, $V = \text{var.}(\eta'\tilde{\beta})$ is monotonic increasing in h, with a point of inflexion at $h = \frac{1}{2}$, for all η. Considering just those η for which $\eta'\beta^* \neq 0$, $\bar{B}$ and λ^* are convex in h; and finally λ is convex in h for all η.

Thus, the essential strengthening of the previous results arising from the replacement of A^{-1} by $h(X'X)^{-1}$ is that $\bar{B}$ and λ^* are now convex in h. This arises directly from the fact that the Curve Décolletage is now a straight line, so that the indeterminacies of the general case are eliminated.

The curves for $Bias^2(\eta'\tilde{\beta})$ in Figures III.IV.4 to III.IV.6 are now replaced by convex functions, and the curves for M.S.E. $(\eta'\tilde{\beta})$ are modified accordingly.

It is clear from Leamer's discussion that when h = 1 the interpretation problem associated with multicollinearity vanishes.[21] This fact highlights the relationship between X and A in $\tilde{\beta}$. Of course, there may still be a weak data problem which will be reflected in the estimation precision. In particular, if the data are perfectly collinear neither $\hat{\beta}$ nor $\tilde{\beta}$ will be obtainable.

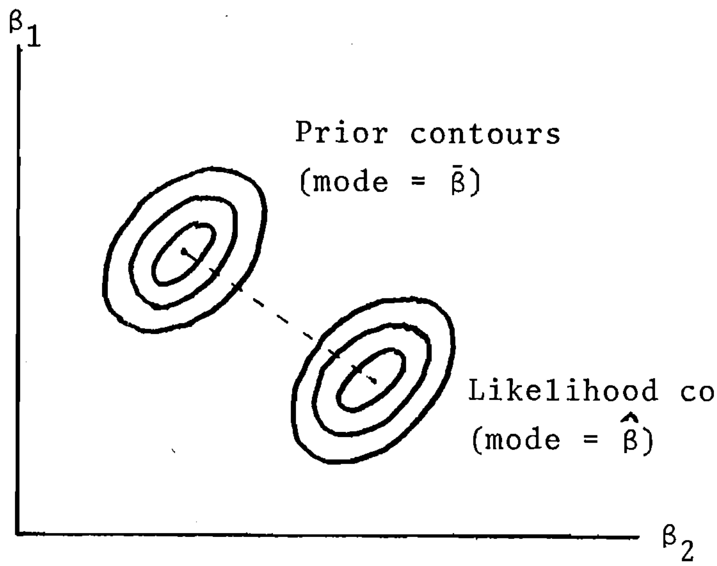


Figure III.V.1: The curve décolletage when $A \propto (X'X)$

VI. A TEST STATISTIC

(1) A Formal Test

In this Section we take a frequentist view and consider the possibility of a formal test of the hypothesis $H_0 : \lambda \leq 1$, vs. $H_1 : \lambda > 1$. Such a test may be feasible if λ can be shown

---

21. Leamer, op. cit., p.377.

to be a parameter in some known distribution. Following Toro-Vizcarrondo and Wallace, a natural possibility is the non-central F distribution.

We emphasize again that if one of $\tilde{\beta}$ or $\hat{\beta}$ is to be selected subsequent to testing the above hypotheses, then attention should be paid to the pre-testing bias problem, and some generalization of the analyses of Bock et al. and of Judge et al. (1973) would be relevant here. However, we do not pursue these possibilities in this thesis.

Consider the two quadratic forms:

$$\text{(i)} \qquad (Q_1/\sigma^2) = \sigma^{-2}\text{S.S.E.}(\hat{\beta})$$

$$= (y/\sigma)'M(y/\sigma),$$

where $M = [I - X(X'X)^{-1}X']$ is idempotent and symmetric.

$$\text{(ii)} \qquad (Q_0/\sigma^2) = \sigma^{-2}\{\text{S.S.E.}(\tilde{\beta}) - \text{S.S.E.}(\hat{\beta})\}$$

$$= \sigma^{-2}(\hat{\beta}-\bar{\beta})'\{A'W'X'XWA\}(\hat{\beta}-\bar{\beta}),$$

where use is made of the relationship

$$\tilde{\beta} = \hat{\beta} - WA(\hat{\beta}-\bar{\beta}).$$

Now, (Q_1/σ^2) is a quadratic form in the random vector (y/σ), this vector being distributed $N[(X\beta/\sigma), I_k]$ so (Q_1/σ^2) is $\chi^2_{(n-k)}$. Further, (Q_0/σ^2) is a quadratic form in the random vector $[(\hat{\beta}-\bar{\beta})/\sigma]$, this vector being distributed $N[(\beta-\bar{\beta})/\sigma, (X'X)^{-1}]$, so (Q_0/σ^2) is non-central $\chi^2_{(k)}$ iff

$$Z = A'W'X'XWA(X'X)^{-1}$$

is idempotent, by Theorem A.10. In general, Z is <u>not</u> idempotent. However, since

$$Z = (I+\phi X'XA_0^{-1})^{-1}(X'X)(I+\phi A_0^{-1}X'X)^{-1}(X'X)^{-1},$$

then

$$\underset{\phi \to 0}{\text{Limit}}(Z) = I_k$$

which is idempotent.

Recall from Part (2) of Section IV that as $\phi \to 0$, the conditional prior covariance matrix for the elements of β approaches the zero matrix. That is, using $\tilde{\beta}$ amounts to imposing the <u>exact</u> (vector) restriction $\beta = \bar{\beta}$, as a way of "estimating" β - the <u>a priori</u> information totally dominates any sample information.

In this strong limiting case, (Q_0/σ^2) is non-central $\chi^2_{(k)}$ with non-centrality parameter

$$\begin{aligned} \rho &= \tfrac{1}{2}\sigma^{-2}\beta^{*\prime}(X'X)\beta^{*} \\ &= \tfrac{1}{2}\underset{\phi \to 0}{\text{Limit}}(\lambda) \\ &= \tfrac{1}{2}\lambda_L\ , \quad \text{say.} \end{aligned} \qquad \text{III.VI.1}$$

Now, if (Q_0/σ^2) and (Q_1/σ^2) are independent, then the <u>observable</u> quantity $\{(n-k)Q_0/kQ_1\}$ is non-central F with k and (n-k) degrees of freedom, in this special limiting case, and the non-centrality parameter is $\rho = \frac{1}{2}\lambda_L$ from III.VI.1. In this case the hypothesis $H_0' : \rho \leq \frac{1}{2}$ is equivalent to $H_0 : \lambda \leq 1$, and is testable. A proof of independence follows, and it does not rely on taking the limiting value of A.

Proposition III.VI.1:

The quadratic forms (Q_0/σ^2) and (Q_1/σ^2) are independently distributed.

Proof: $(Q_0/\sigma^2) \;=\; m'(A'W'X'XWA)m,$

where $\quad m \;=\; (\hat{\beta} - \bar{\beta})/\sigma.$

Now, applying Theorem A.1 twice, $(A'W'X'XWA)$ is p.d.s. and hence it is non-singular, by Theorem A.4 . Further, by Theorem A.12, there exists a $(k \times k)$ matrix L of rank k such that $A'W'X'XWA = LL'$.

Also, $(Q_1/\sigma^2) = z'Mz$,

where $z = (y/\sigma)$.

Now, rank(M) = (n-k), so there exists an $[n \times (n-k)]$ matrix, N, with rank (n-k), such that $M = NN'$, by Theorem A.12, since M is symmetric.

Let

$$
\begin{aligned}
C &= \text{cov.}\{L'm, z'N\} \\
&= \mathcal{E}\{[L'm - \mathcal{E}(L'm)][z'N - \mathcal{E}(z'N)]\} \\
&= \sigma^{-2}\{\mathcal{E}(L'\hat{\beta}y'N) - L'\beta\beta'X'N\}
\end{aligned}
$$

Now,

$$\mathcal{E}(L'\hat{\beta}y'N) = \mathcal{E}\{L'(X'X)^{-1}X'yy'N\}$$

$$= L'\beta\beta'X'N + \sigma^2 L'(X'X)^{-1}X'N$$

So,

$$C = L'(X'X)^{-1}X'N$$

Now,

$$CN' = L'(X'X)^{-1}X'NN'$$

$$= L'(X'X)^{-1}X'M$$

$$= 0.$$

Thus, $CN'N = 0$.

But, $(N'N)^{-1}$ exists, since N has full column rank, so

$$C(N'N)(N'N)^{-1} = C$$

$$= 0.$$

Both m and z are normally distributed, so by Theorem A-11, $L'm$ and $z'N$ are independent. Thus,

$$(Q_0/\sigma^2) = m'(A'W'X'XWA)m$$

$$= (L'm)'(L'm),$$

and,

$$(Q_1/\sigma^2) = z'Mz = (z'N)(z'N)'$$

are also independently distributed. <u>Q.E.D.</u>

This result holds for <u>all</u> A, and in particular in the limiting case as $\phi \to 0$, for then $(A'W'X'XWA)$ collapses to $(X'X)^{-1}$ which is symmetric and of full column rank, k.

Thus, in the limit as $\phi \to 0$ the test procedure derived by Toro-Vizcarrondo and Wallace, and the tables developed by Wallace and Toro-Vizcarrondo (1969), may be applied to our problem, making use of the fact that the observable quantity $[(n-k)Q_0/kQ_1]$ is non-central F with k and (n-k) degrees of freedom and non-centrality parameter $\rho = \frac{1}{2}\lambda_L$.

Indeed, this limiting case is a <u>very strong</u> special example of the general linear restrictions problem considered by Toro-Vizcarrondo and Wallace.[22] They consider restrictions of the form

$$H'\beta = h, \qquad \text{III.VI.2}$$

where H' is $(m \times k)$ of rank $m \leq k$, h is $(m \times 1)$, and both H and h are fixed and known. Our limiting case as $\phi \to 0$ is just a special case of III.VI.2, with exact restrictions on <u>all</u> of the elements of β, and with $m = k$, $H = I_k$, and $h = \bar{\beta}$. In this degenerate case, $\tilde{\beta} = \bar{\beta}$, and so it is not surprising that we are able to utilize the non-central F distribution as $\phi \to 0$.

Further, note that if the exact limiting restriction $\beta = \bar{\beta}$ is <u>true</u>, then $\rho = 0$, and $[(n-k)Q_0/kQ_1]$ takes a <u>central</u> F distribution.

However, although this relationship between our problem and that of Toro-Vizcarrondo and Wallace is interesting, it refers only to a degenerate situation. We have been unable to establish a formal test for H_0 in general[23], and so we consider

22. See Judge <u>et al.</u>, <u>op. cit.</u>, for an extension to stochastic restrictions which is closely related to our analysis, but is based on the <u>weak</u> M.S.E. criterion.
23. Griffiths, <u>op. cit.</u>, reports a similar failure in his case.

the possibility of a workable <u>ad hoc</u> test procedure.

(2) <u>An Ad Hoc Test</u>

The true λ is unobservable, but a simple observable substitute[24] is

$$\hat{\lambda} = \hat{\sigma}^{-2}(\hat{\beta}-\bar{\beta})'\{(X'X)^{-1} + 2A^{-1}\}^{-1}(\hat{\beta}-\bar{\beta}) \qquad \text{III.VI.3}$$

where $\hat{\beta}$ is defined in III.II.2, and

$$\hat{\sigma}^2 = (y-X\hat{\beta})'(y-X\hat{\beta})/(n-k)$$

is the O.L.S. estimator of σ^2.

One might consider using $\hat{\lambda}$ in place of λ. That is, after obtaining $\hat{\beta}$ and $\hat{\sigma}^2$, $\hat{\lambda}$ could be computed and if $\hat{\lambda} \leq 1$ then $\tilde{\beta}$ might be inferred to have smaller matrix M.S.E. than $\hat{\beta}$. Again, we emphasize that we are ignoring the important pre-testing bias problem in this context. However, the use of $\hat{\lambda}$ is appealing for several reasons: under the assumptions for III.II.1, $\hat{\beta}$ and $\hat{\sigma}^2$ are Maximum Likelihood estimates; they are easily computed and depend only on sample information; and finally $\hat{\beta}$ is likely to be required in any case, either because $\hat{\lambda} > 1$, or because one will wish to use the fact that $\hat{\beta}$ is based on the diffuse "reference prior". Thus, working with $\hat{\lambda}$ involves certain computational efficiencies.

Of course, these considerations alone are insufficient justification for a direct substitution of $\hat{\lambda}$ for λ. The performance of the test procedure based on $\hat{\lambda}$ is most important,

---

24. See also Allen (1971); McCallum (1970), p.112; and Zellner and Vandaele, <u>op. cit.</u>, p.17.

as are the properties of $\hat{\lambda}$ as an estimator of λ.

(a) Properties of the Ad Hoc Test First, we note an interesting general limiting property of $\hat{\lambda}$:

$$\underset{\phi \to 0}{\text{Limit}}(\hat{\lambda}) = \hat{\sigma}^{-2}(\hat{\beta}-\bar{\beta})'(X'X)(\hat{\beta}-\bar{\beta})$$

$$= \hat{\lambda}_L, \quad \text{say.}$$

Now, $\hat{\lambda}_L$ is related to a well-known statistic:

$$\mathcal{F} = (\hat{\beta}-\bar{\beta})'(X'X)(\hat{\beta}-\bar{\beta})/k\hat{\sigma}^2$$

which is F-distributed with k and (n-k) degrees of freedom.[25] Thus, $(\hat{\lambda}_L/k)$ may be used to test the hypothesis that $\beta = \bar{\beta}$, using the properties of the F distribution. There is an obvious relationship between testing this hypothesis and testing to see if $\hat{\lambda}_L \leq 1$. For a given significance level, a sufficiently small value of $\mathcal{F}$ implies non-rejection of the hypothesis that $\beta = \bar{\beta}$, while a sufficiently small value of $\hat{\lambda}_L$ leads to a preference for $\tilde{\beta}$ over $\hat{\beta}$ (in terms of our strong M.S.E. criterion). These two facts are consistent, as is clear from the discussion in Part (2) of Section IV.

We turn now to the properties of $\hat{\lambda}$ as an estimator of λ.

Proposition III.VI.2:

$\hat{\lambda}$ is a consistent estimator of λ.

Proof: $\underset{n \to \infty}{\text{plim}}(\hat{\sigma}^2) = \sigma^2$,

and $\hat{\sigma}^{-2}$ is continuous in $\hat{\sigma}^2$, so by Theorem A.13,

25. See Johnston (1972), p.152.

$$\operatorname*{plim}_{n \to \infty}(\hat{\sigma}^{-2}) = \sigma^{-2}.$$

Let $J = \{(X'X)^{-1} + 2A^{-1}\}^{-1}$. Then, since $\operatorname*{plim}_{n \to \infty}(\hat{\beta}) = \beta$, and since both $\hat{\beta}'J\hat{\beta}$ and $\bar{\beta}'J\hat{\beta}$ are continuous in $\hat{\beta}$, it follows from Theorem A.13 that

$$\operatorname*{plim}_{n \to \infty}(\bar{\beta}'J\hat{\beta}) = \bar{\beta}'J\beta$$

and

$$\operatorname*{plim}_{n \to \infty}(\hat{\beta}'J\hat{\beta}) = \beta'J\beta.$$

Now,

$$\hat{\lambda} = \hat{\sigma}^{-2}\{\hat{\beta}'J\hat{\beta} - 2\bar{\beta}'J\hat{\beta} + \bar{\beta}'J\bar{\beta}\}$$

So,[26]

$$\operatorname*{plim}_{n \to \infty}(\hat{\lambda}) = \sigma^{-2}\{\beta'J\beta - 2\bar{\beta}'J\beta + \bar{\beta}'J\bar{\beta}\}$$

$$= \lambda.$$

Q.E.D.

Thus, there is at least a large-sample justification for replacing $\hat{\lambda}$ by λ. However, in most practical situations the small-sample properties of $\hat{\lambda}$ are likely to be of more concern. First, a preliminary result must be established.

Let:

$$Q_2 = (y-X\hat{\beta})'(y-X\hat{\beta})$$

$$= y'My,$$

where $M = [I-X(X'X)^{-1}X']$.

Thus, $\hat{\sigma}^2 = Q_2/(n-k)$. Further, let

$$Q_3 = (\hat{\beta}-\bar{\beta})'J(\hat{\beta}-\bar{\beta}),$$

26. See Christ (1966), pp.375-376.

where J is defined as above. Now, Q_2 is a quadratic form in the random vector y, the latter being distributed $N[X\beta, \sigma^2 I_k]$, and Q_3 is a quadratic form in the random vector $(\hat{\beta}-\bar{\beta})$, the latter being distributed $N[(\beta-\bar{\beta}), \sigma^2(X'X)^{-1}]$.

<u>Proposition III.VI.3</u>:

Q_2 and Q_3 are independently distributed.

<u>Proof</u>: J is symmetric, so there exists a $(k \times k)$ matrix R of rank k such that $J = RR'$, by Theorem A.12. Further, M is symmetric with rank (n-k), so by Theorem A.12 there exists an $[n \times (n-k)]$ matrix N of rank (n-k) such that $M = NN'$.

Let

$$C = \text{cov.}\{R'(\hat{\beta}-\bar{\beta}), y'N\}$$

Then,

$$\begin{aligned} C &= \mathcal{E}\{[R'(\hat{\beta}-\bar{\beta}) - R'(\beta-\bar{\beta})][y'N - \beta'X'N] \\ &= \mathcal{E}\{R'(\hat{\beta}-\beta)\varepsilon'N\} \\ &= \sigma^2(R'(X'X)^{-1}X'N) \end{aligned}$$

Thus,

$$\begin{aligned} CN' &= \sigma^2 R(X'X)^{-1}X'M \\ &= 0, \end{aligned}$$

from the definition of M.

So,

$$CN'N = 0,$$

and since N has full column rank, $(N'N)^{-1}$ exists, and,

$$C(N'N)(N'N)^{-1} = C = 0.$$

Applying Theorem A.11, $R'(\hat{\beta}-\bar{\beta})$ and $y'N$ are independently distributed, and thus so are

$$Q_2 = (y'N)(y'N)'$$

and

$$Q_3 = [R'(\hat{\beta}-\bar{\beta})]'[R'(\hat{\beta}-\bar{\beta})].$$ <u>Q.E.D.</u>

Since $\hat{\sigma}^2 = Q_2/(n-k)$, it follows immediately that $\hat{\sigma}^2$ and Q_3 are independently distributed. Now, consider the bias of $\hat{\lambda}$ as an estimator of λ.

<u>Proposition III.VI.4</u>:

$\hat{\lambda}$ is upward-biased in small samples, and $\text{Bias}(\hat{\lambda}) > \text{tr.}\{[I + 2X'XA^{-1}]^{-1}\}$.

<u>Proof</u>: $\hat{\lambda} = (Q_3/\hat{\sigma}^2)$, where Q_3 and $\hat{\sigma}^2$ are independently distributed. Further, $\hat{\sigma}^2 > 0$; and since J is p.d., $Q_3 > 0$. Thus, applying Theorem A.17,

$$\mathcal{E}(\hat{\lambda}) > \mathcal{E}(Q_3)/\mathcal{E}(\hat{\sigma}^2).$$

As is well-known, $\mathcal{E}(\hat{\sigma}^2) = \sigma^2$. Further, from Theorem A.18,

$$\mathcal{E}(Q_3) = (\beta-\bar{\beta})'J(\beta-\bar{\beta}) + \sigma^2\text{tr.}\{[I+2X'XA^{-1}]^{-1}\}$$

Thus,

$$\mathcal{E}(\hat{\lambda}) > \lambda + \text{tr.}\{[I+2X'XA^{-1}]^{-1}\}.$$

Now, Q_3 is a convex function[27] of $(\hat{\beta}-\bar{\beta})$, since J is p.d.s.. By Theorem A.14, since $\mathcal{E}(\hat{\beta}-\bar{\beta}) = (\beta-\bar{\beta})$,

$$\mathcal{E}(Q_3) > (\beta-\bar{\beta})'J(\beta-\bar{\beta})$$

so that $L = \text{tr.}\{[I + 2X'XA^{-1}]^{-1}\} > 0$.

Thus,

$$\begin{aligned}\text{Bias}(\hat{\lambda}) &= \mathcal{E}(\hat{\lambda}) - \lambda \\ &> L \\ &> 0\end{aligned}$$

<u>Q.E.D.</u>

Note that in the special case where σ is <u>known</u>, then

$$\begin{aligned}\mathcal{E}(\hat{\lambda}) &= \sigma^{-2}\mathcal{E}(Q_3) \\ &= \lambda + L\end{aligned}$$

so that $\text{Bias}(\hat{\lambda}) = L$, exactly in this case.

An important feature of L is that it depends on the particular sample values, as well as on A. Although L is independent of β, σ^2 and $\bar{\beta}$, it is quite possible that $\text{Bias}(\hat{\lambda})$ itself could depend on these other parameters.

As n increases, the diagonal elements of $(X'XA^{-1})$ increase, so that $\underset{n \to \infty}{\text{Limit}}(L) = 0$. This accords with the fact that $\hat{\lambda}$ is consistent and hence asymptotically unbiased. Further,

$$L = \phi^{-1} \text{ tr.}\{[\phi^{-1}I + 2X'XA_0^{-1}]^{-1}\},$$

27. See Hadley (1964), pp.84-85.

as $\underset{\phi \to \infty}{\text{Limit}}(L) = 0$. That is, as $\tilde{\beta} \to \hat{\beta}$ the lower bound on Bias(L) vanishes.

Since a lower bound (not necessarily the greatest lower bound) for Bias$(\hat{\lambda})$ is known exactly and is a function of observable quantities, it would be advisable to replace $\hat{\lambda}$ by $\tilde{\lambda} = (\hat{\lambda}\text{-}L)$ for the purposes of testing to see if $\tilde{\beta}$ is "preferred" to $\hat{\beta}$. Clearly, $\tilde{\lambda}$ is a consistent estimator of λ and will have smaller bias than has $\tilde{\lambda}$ in small samples.

However, the magnitudes of the biases and the other small-sample properties of $\hat{\lambda}$ and $\tilde{\lambda}$ are not obtainable analytically. Some of these are investigated in a limited Monte Carlo experiment.

(b) A Monte Carlo Experiment The computational burden of the experiment was limited by restricting attention to the simple model:

$$y_t = \beta x_t + \varepsilon_t \; ; \qquad t = 1,2,\ldots,n. \qquad \text{III.IV.4}$$

where now (and henceforth in this Chapter) β and $\bar{\beta}$ are scalar parameters, and A^{-1} is replaced by the positive scalar, a.

Then,

$$\lambda = [(\beta-\bar{\beta})^2 \sum_t x_t^{\,2}] / [\sigma^2 (1 + 2a\sum_t x_t^{\,2})] \qquad \text{III.VI.5}$$

$$\hat{\lambda} = [(\hat{\beta}-\bar{\beta})^2 \sum_t x_t^{\,2}] / [\hat{\sigma}^2 (1 + 2a\sum_t x_t^{\,2})]$$

and,

$$\tilde{\lambda} = \hat{\lambda} - 1/(1 + 2a\sum_t x_t^{\,2})$$

where:

$$\hat{\beta} = (\sum_t x_t y_t)/(\sum_t x_t^{\,2})$$

$$\hat{\sigma}^2 = [\sum_t (y_t - \hat{\beta}x_t)^2]/(n-1).$$

Once the values of β, σ^2, $\bar{\beta}$, a, n and $\{x_t\}$ were fixed, one hundred samples (of size n) of ε_t were drawn[28] from $N(0,\sigma^2)$; the corresponding $\{y_t\}$ series were constructed from III.VI.4; and one hundred values of $\hat{\lambda}$ were computed for comparison with λ. These $\hat{\lambda}$ values were also used to obtain:

(i) $\bar{\hat{\lambda}} = (\sum_i \hat{\lambda}_i)/100$

(ii) $\text{Bias}(\hat{\lambda}) = (\bar{\hat{\lambda}} - \lambda)$

(iii) $\text{var.}(\hat{\lambda}) = \sum_i (\hat{\lambda}_i - \bar{\hat{\lambda}})^2/100$

(iv) $\text{M.S.E.}(\hat{\lambda}) = \text{var.}(\hat{\lambda}) + \text{Bias}^2(\hat{\lambda})$

(v) Π = the frequency with which $\hat{\lambda}_i > 1$.

Attention was focused on $\hat{\lambda}$ rather than $\tilde{\lambda}$ since for the values of a and $\{x_t\}$ investigated, $L = 1/(1+2a\sum_t x_t^2)$ was negligible[29] even for n = 10. Thus, in this particular case knowledge of the value of L was of no practical help in reducing the bias, *etc.*, of $\hat{\lambda}$. However, from III.VI.5 it is clear that situations could arise in practice where L is of a substantial magnitude, in which case $\tilde{\lambda}$ would clearly be more useful than $\hat{\lambda}$. Since L is observable, this poses no practical difficulty.

---

28. The normally distributed disturbances are generated within the computer program written for the experiment. Twenty four deviates, u_i, from a uniform [0,1] distribution are generated by the RANDOM intrinsic on the Burroughs 6718 machine, and are converted to a single $N(0,\sigma^2)$ deviate by the transformation
$$\varepsilon = (\sigma/\sqrt{2})\,[(\sum_1^{24} u_i) - 12].$$
See Hamming(1962), p.389.
29. A limited number of tests were made with $\tilde{\lambda}$ in place of $\hat{\lambda}$. The results obtained generally differed by less than 0.1% from those for the latter.

All combinations of the following were studied:

(i) The $\{x_t\}$ series[30] was fixed throughout and comprised two parts - a growth component of base 10.0 and 2% growth per observation; and a random component from N(0,1).

(ii) n = 10; 30.

(iii) σ^2 = 1.0; 2.0

(iv) β = 1.0; 2.0; 3.0.

(v) $\bar{\beta}$ takes four values for each value of β in (iv):

β = 1.0 : $\bar{\beta}$ = 0.06; 0.10; 0.50; 0.80

β = 2.0 : $\bar{\beta}$ = 0.12; 0.20; 1.00; 1.60

β = 3.0 : $\bar{\beta}$ = 0.18; 0.30; 1.50; 2.40.

(vi) λ = 0.4 (0.1) 0.8; 0.9 (0.02) 1.1; 1.2 (0.1) 1.6.

(vii) a was set as the dependent variable and adjusted to ensure that the desired value of λ was achieved in any particular situation.

The values chosen for (i), (ii) and (iii) imply realistic "Hybrid R^2" values for III.VI.4, these values being obtained from,

$$R^2 = [\beta^2 \text{var.}(x)] / [\beta^2 \text{var.}(x) + \sigma^2]$$

where var.(x) is the <u>sample</u> variance of $\{x_t\}$. These R^2 values appear in Table III.VI.1.

---

30. This series is also used in Section IV of Chapter VI.

TABLE III.VI.1

Hybrid R^2 Values

σ^2	n	β 1.0	2.0	3.0
1	10	0.628	0.871	0.938
1	30	0.857	0.960	0.982
2	10	0.458	0.772	0.884
2	30	0.750	0.923	0.964

Two factors generalize the results obtained and limit the situations that need to be studied. First, λ is symmetric in $(\bar{\beta}-\beta)$, so _ceteris paribus_ the results obtained with a range of values of $\bar{\beta} > \beta$ apply equally for the symmetric[31] values of $\bar{\beta} < \beta$.

Secondly, there is no need to consider situations which differ from one already studied only in that β, $\bar{\beta}$ and σ are all scaled by the same multiplicative factor. Note that if these parameters are scaled multiplicatively, then so are $\hat{\beta}$ and $\hat{\sigma}$:

(i) $\hat{\beta} = \beta + (\sum_t x_t \varepsilon_t)/(\sum_t x_t^2)$. If σ is scaled by ∇ then so is $\sum_t x_t \varepsilon_t$. If β is also scaled by ∇, then so is $\hat{\beta}$.

(ii) $\hat{\sigma}^2 = [\sum_t \varepsilon_t^2 - (\sum_t x_t \varepsilon_t)^2/(\sum_t x_t^2)]/(n-1)$. If σ is scaled by ∇, then $\sum_t \varepsilon_t^2$, $(\sum_t x_t \varepsilon_t)^2$ and $\hat{\sigma}^2$ are scaled by ∇^2, so $\hat{\sigma}$ is scaled by ∇.

31. For example, if $\beta = 1.0$, the results obtained with $\bar{\beta} = 0.9$ apply equally when $\bar{\beta} = 1.1$, _ceteris paribus_.

Thus, since

$$\hat{\lambda} = \lambda\{[(\hat{\beta}-\bar{\beta})/\hat{\sigma}]/[(\beta-\bar{\beta})/\sigma]\}^2,$$

if β, $\bar{\beta}$ and σ are scaled multiplicatively by ∇, $\hat{\lambda}$ just varies with λ.

Finally, for the simple model III.VI.4, it is meaningful to consider the "relative efficiency" of $\hat{\beta}$ to $\tilde{\beta}$:

$$e = \text{M.S.E.}(\tilde{\beta})/\text{M.S.E.}(\hat{\beta}),$$

and it may be useful to compare e with λ. They are related by

$$e = \lambda + (1-\lambda)h^2,$$

where

$$h = [(a\sum_t x_t^2)/(1 + a\sum_t x_t^2)] < 1$$

Although e and λ differ in general, it is clear that $e = 1$ if $\lambda = 1$; and that $e = h^2 (< 1)$ if $\lambda = 0$. Further, e is a monotonic increasing function of λ, since

$$(\partial e/\partial\lambda) = (1-h^2) > 0,$$

since $h < 1$.

The relationship between e and $\hat{\lambda}$ is

$$e = \hat{\lambda}g^2 + (1 - \hat{\lambda}g^2)h^2,$$

where

$$g = \{[(\beta-\bar{\beta})/\sigma]/[(\hat{\beta}-\bar{\beta})/\hat{\sigma}]\}.$$

Thus, $e = g^2 + (1-g^2)h^2$ ($\neq 1$) when $\hat{\lambda} = 1$; and $e = h^2$ (< 1) when $\hat{\lambda} = 0$. Of course, in large samples $\hat{\lambda} \to \lambda$, and the relationships between e and λ apply.[32] Further, e is monotonic increasing in $\hat{\lambda}$, for all $\bar{\beta} \neq \beta$, since

$$(\partial e/\partial\hat{\lambda}) = g^2(1-h^2) > 0$$

if $g \neq 0$, since $h < 1$.

(c) Results

The results of the Monte Carlo experiment[33] appear in Tables III.VI.2 to III.VI.13. The results obtained are insensitive[34] to the choice of $\bar{\beta}$, *ceteris paribus*, so the tabulated figures are averages across the four relevant values of $\bar{\beta}$ in each case.

In the limited range of situations tested, $\hat{\lambda}$ is found to be upward-biased by about 30% when n = 10, and about 8% when n = 30. For a particular sample size this bias is quite insensitive to changes in the various other factors. Again, this need not necessarily be the case in general.

Further, $\hat{\lambda}$ has a rather large sampling variance - about 71% when n = 10, and about 14% when n = 30. The decrease in both bias and variance as n increases reflects the consistency of $\hat{\lambda}$, but this estimator (and $\tilde{\lambda}$) should be treated with caution in samples of less than thirty observations.

The values of Π in Tables III.VI.2 to III.VI.13 are used to estimate power curves for the test procedure:

---

32. In large samples, $\hat{\beta} \to \beta$ and $\hat{\sigma} \to \sigma$, so $g \to 1$ and then $e = 1$ if $\hat{\lambda} = 1$.
33. In the tables of results, Bias($\hat{\lambda}$), var.($\hat{\lambda}$) and M.S.E.($\hat{\lambda}$) are all expressed as percentages of the true λ.
34. Of course, in general this need not be the case.

TABLE III.VI.2

(β=1.0; σ^2=1.0; n=10)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE ($\hat{\lambda}$)
0.40	0.1275	0.591	47.780	69.452	80.333
0.50	0.1750	0.669	33.940	58.101	65.945
0.60	0.2125	0.768	28.000	50.225	56.984
0.70	0.3475	0.984	40.560	78.822	88.668
0.80	0.3675	1.121	40.150	38.761	42.542
0.90	0.4775	1.276	41.800	95.965	108.099
0.92	0.5025	1.318	43.240	104.962	120.990
0.94	0.5200	1.317	40.130	71.252	81.387
0.96	0.5125	1.282	33.590	51.150	58.009
0.98	0.5775	1.361	38.890	58.854	67.008
1.00	0.5725	1.385	38.490	58.968	69.978
1.02	0.5525	1.449	42.020	82.737	94.287
1.04	0.5875	1.487	42.960	65.269	73.647
1.06	0.5500	1.443	36.160	69.885	77.212
1.08	0.5600	1.527	41.390	78.879	87.333
1.10	0.6275	1.576	43.240	57.367	64.548
1.20	0.6250	1.550	29.210	54.973	58.810
1.30	0.7400	1.803	38.680	95.927	107.996
1.40	0.7575	1.990	42.160	72.909	81.757
1.50	0.7800	1.948	29.850	94.578	107.313
1.60	0.8150	2.124	32.760	74.201	85.261
		Average:	38.330	70.630	79.910

* True relative efficiency, e, ranges from 0.760 to 1.536.

TABLE III.VI.3

(β=1.0; σ^2=1.0; n=30)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0075	0.417	4.280	7.265	7.398
0.50	0.0475	0.554	10.880	11.474	12.117
0.60	0.0700	0.629	4.880	9.685	9.872
0.70	0.1600	0.759	8.410	9.969	12.202
0.80	0.3050	0.895	11.850	16.375	17.898
0.90	0.3850	0.969	7.640	15.635	16.300
0.92	0.4075	0.997	8.380	18.929	19.695
0.94	0.4450	1.056	12.320	18.948	20.851
0.96	0.4625	1.040	8.380	20.720	21.697
0.98	0.5075	1.084	10.580	22.019	23.484
1.00	0.5500	1.128	12.800	21.105	23.117
1.02	0.5650	1.159	13.610	21.245	24.136
1.04	0.5975	1.138	9.410	17.428	18.404
1.06	0.6050	1.162	9.640	21.621	24.001
1.08	0.6775	1.242	15.050	20.705	23.404
1.10	0.6375	1.165	5.940	17.923	18.366
1.20	0.7375	1.324	10.360	20.108	23.012
1.30	0.7950	1.397	7.440	20.678	23.315
1.40	0.8650	1.545	10.330	21.799	24.215
1.50	0.8900	1.597	6.450	20.082	22.986
1.60	0.9150	1.745	9.040	21.514	23.982
		Average:	9.410	17.868	19.545

* True relative efficiency, e, ranges from 0.938 to 1.214.

TABLE III.VI.4

(β=2.0; σ^2=1.0; n=10)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0725	0.509	27.180	26.014	29.039
0.50	0.1600	0.679	35.760	40.436	47.299
0.60	0.2250	0.801	33.550	51.608	58.478
0.70	0.3550	0.953	36.110	55.719	56.296
0.80	0.4250	1.095	36.880	77.152	88.822
0.90	0.4625	1.156	28.470	76.287	83.646
0.92	0.5225	1.270	38.010	86.718	90.113
0.94	0.4950	1.201	27.790	66.265	74.349
0.96	0.5375	1.230	28.100	62.673	70.918
0.98	0.5275	1.240	26.580	64.904	72.119
1.00	0.5725	1.276	27.600	83.276	91.101
1.02	0.5600	1.323	29.720	96.611	105.661
1.04	0.5600	1.370	31.750	92.023	105.015
1.06	0.5925	1.458	37.580	60.606	70.522
1.08	0.6425	1.461	35.300	90.618	103.261
1.10	0.5950	1.398	27.050	96.251	105.592
1.20	0.6725	1.510	25.830	94.255	105.311
1.30	0.7525	1.694	30.350	90.275	103.014
1.40	0.8100	1.886	34.710	86.421	90.001
1.50	0.8225	1.876	25.080	96.739	105.842
1.60	0.8500	1.975	23.430	74.312	81.421
		Average:	30.800	74.722	82.755

* True relative efficiency, e, ranges from 0.929 to 1.240.

TABLE III.VI.5

(β=2.0; σ^2=1.0; n=30)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0025	0.430	7.430	5.028	5.268
0.50	0.0225	0.545	9.020	6.424	7.711
0.60	0.0550	0.646	7.620	6.184	6.565
0.70	0.1250	0.747	6.690	7.629	8.012
0.80	0.2000	0.832	4.040	7.616	7.829
0.90	0.3975	0.959	6.580	9.091	9.631
0.92	0.4375	0.998	8.470	10.365	11.068
0.94	0.4500	1.016	8.120	9.947	10.603
0.96	0.4650	1.058	10.210	11.864	12.728
0.98	0.4775	1.045	6.630	10.734	11.173
1.00	0.5200	1.064	6.350	11.526	11.968
1.02	0.5450	1.088	6.620	11.611	12.071
1.04	0.5725	1.125	8.200	12.242	12.976
1.06	0.5575	1.127	6.350	12.403	12.991
1.08	0.6175	1.136	5.170	10.695	11.098
1.10	0.6625	1.191	8.250	14.427	15.223
1.20	0.7500	1.285	7.080	13.440	14.112
1.30	0.8075	1.356	4.350	14.006	14.432
1.40	0.8900	1.510	7.860	14.930	15.805
1.50	0.9425	1.648	9.850	11.133	12.159
1.60	0.9500	1.704	6.520	14.313	14.895
		Average:	7.210	10.744	11.348

* True relative efficiency, e, ranges from 0.984 to 1.062.

TABLE III.VI.6

(β=3.0; σ^2=1.0; n=10)*

λ	Π	$\bar{\bar{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0725	0.551	37.750	58.571	61.497
0.50	0.1375	0.653	30.660	25.189	29.905
0.60	0.1925	0.763	27.200	43.569	48.372
0.70	0.2975	0.875	25.000	50.896	57.283
0.80	0.3850	1.011	26.380	45.757	51.478
0.90	0.4725	1.191	32.330	87.710	97.614
0.92	0.5275	1.195	29.890	53.387	62.984
0.94	0.4925	1.189	26.490	54.758	62.373
0.96	0.4900	1.235	28.650	69.852	79.280
0.98	0.5800	1.308	33.470	67.608	78.820
1.00	0.5425	1.331	33.100	77.417	88.686
1.02	0.5275	1.336	30.980	98.989	109.012
1.04	0.5600	1.337	28.560	72.275	81.077
1.06	0.5900	1.349	27.260	68.657	77.056
1.08	0.6550	1.430	32.410	92.398	104.937
1.10	0.6375	1.404	27.640	60.566	69.300
1.20	0.7325	1.596	33.000	77.581	90.969
1.30	0.7300	1.631	25.480	84.476	93.034
1.40	0.7825	1.701	21.510	61.260	68.077
1.50	0.8325	1.918	27.890	60.564	67.514
1.60	0.8725	2.073	29.560	59.424	66.021
		Average:	29.300	65.281	73.585

* True relative efficiency, e, ranges from 0.999 to 1.121.

TABLE III.VI.7

(β=3.0; σ^2=1.0; n=30)*

λ	Π	$\bar{\bar{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0025	0.438	9.500	3.936	4.305
0.50	0.0050	0.520	4.000	4.405	4.496
0.60	0.0400	0.642	7.030	5.803	6.110
0.70	0.1125	0.762	8.900	8.488	9.089
0.80	0.2625	0.870	8.750	8.480	9.183
0.90	0.3800	0.983	9.220	11.197	11.993
0.92	0.4325	0.999	8.590	8.878	9.672
0.94	0.4450	1.021	8.620	10.512	11.427
0.96	0.4325	1.015	5.730	9.294	9.385
0.98	0.5175	1.059	8.060	9.242	9.909
1.00	0.5700	1.091	9.100	9.440	10.313
1.02	0.5450	1.080	5.880	8.464	8.834
1.04	0.5975	1.138	9.420	10.763	11.773
1.06	0.5975	1.124	6.040	10.015	10.448
1.08	0.6325	1.149	6.390	10.756	11.221
1.10	0.7025	1.199	9.000	9.744	10.784
1.20	0.7675	1.268	5.670	11.451	11.986
1.30	0.8775	1.423	9.470	12.662	13.898
1.40	0.9350	1.512	8.000	14.624	15.541
1.50	0.9325	1.579	5.290	12.898	13.340
1.60	0.9725	1.694	5.860	14.440	15.025
		Average:	7.550	9.785	10.416

* True relative efficiency, e, ranges from 0.993 to 1.029.

TABLE III.VI.8

(β=1.0; σ^2=2.0; n=10)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0875	0.572	42.880	74.902	82.498
0.50	0.1200	0.647	29.380	47.128	53.049
0.60	0.2225	0.776	29.280	53.234	58.690
0.70	0.3025	0.945	35.060	63.309	72.657
0.80	0.4300	1.038	29.700	50.574	57.664
0.90	0.4775	1.171	30.110	62.928	71.407
0.92	0.5425	1.272	38.270	95.635	110.099
0.94	0.5275	1.266	34.700	91.698	103.862
0.96	0.5075	1.246	29.750	83.774	93.938
0.98	0.5200	1.310	33.690	107.472	119.080
1.00	0.5975	1.338	33.800	76.368	88.230
1.02	0.5800	1.317	29.160	78.723	87.505
1.04	0.5825	1.329	27.810	56.082	64.944
1.06	0.5300	1.301	22.760	81.038	86.541
1.08	0.6025	1.326	22.730	48.471	54.111
1.10	0.5875	1.469	33.580	98.699	112.045
1.20	0.6525	1.581	31.770	67.275	74.701
1.30	0.7375	1.669	28.370	84.326	94.803
1.40	0.8100	1.898	35.610	65.201	73.022
1.50	0.8125	1.961	30.710	72.013	80.049
1.60	0.8425	2.251	40.670	56.215	65.123
		Average:	31.890	72.146	81.144

* True relative efficiency, e, ranges from 0.613 to 1.600.

TABLE III.VI.9

(β=1.0; σ^2=2.0; n=30)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0250	0.436	9.050	12.221	12.352
0.50	0.0700	0.569	13.720	15.836	17.082
0.60	0.1050	0.659	9.800	13.100	13.746
0.70	0.1850	0.772	10.360	17.352	18.580
0.80	0.2450	0.845	5.660	17.475	17.832
0.90	0.4025	0.977	8.600	20.196	20.904
0.92	0.4325	1.027	11.620	24.808	26.251
0.94	0.4275	1.001	6.500	24.201	24.724
0.96	0.5025	1.086	13.140	27.903	30.247
0.98	0.4975	1.106	12.860	28.202	30.826
1.00	0.5357	1.116	11.600	23.262	24.916
1.02	0.5025	1.125	10.300	33.589	34.791
1.04	0.5625	1.153	10.890	26.904	28.304
1.06	0.5475	1.173	10.640	40.044	42.348
1.08	0.6025	1.205	11.570	28.702	31.248
1.10	0.6400	1.255	14.070	33.385	36.689
1.20	0.7150	1.332	10.980	33.025	35.458
1.30	0.7400	1.368	5.230	35.436	36.153
1.40	0.8200	1.556	11.110	35.317	37.402
1.50	0.8650	1.665	11.000	27.247	28.561
1.60	0.8900	1.773	10.830	34.718	36.014
		Average:	10.450	26.330	27.830

* True relative efficiency, e, ranges from 0.882 to 1.354.

TABLE III.VI.10

(β=2.0; σ^2=2.0; n=10)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0825	0.518	29.600	23.375	26.444
0.50	0.1275	0.677	35.380	41.895	46.847
0.60	0.2425	0.763	27.150	34.428	37.857
0.70	0.3225	0.905	29.310	39.526	44.584
0.80	0.3875	1.124	40.540	70.490	81.547
0.90	0.4825	1.190	32.230	51.471	58.963
0.92	0.4675	1.293	40.550	64.861	69.201
0.94	0.5275	1.265	34.530	85.275	95.505
0.96	0.5150	1.301	35.490	112.968	123.929
0.98	0.5175	1.266	29.140	50.566	56.600
1.00	0.5425	1.281	28.080	50.882	59.495
1.02	0.5275	1.294	26.880	62.968	70.208
1.04	0.5725	1.332	28.090	48.363	53.639
1.06	0.5900	1.408	32.780	76.174	84.420
1.08	0.6225	1.387	28.460	57.316	62.874
1.10	0.6525	1.518	37.960	62.404	71.317
1.20	0.6750	1.646	37.130	62.650	69.468
1.30	0.7375	1.843	41.770	119.032	141.326
1.40	0.8025	1.920	37.160	116.274	130.451
1.50	0.7950	2.048	36.500	93.359	108.244
1.60	0.8400	2.232	39.510	115.447	131.611
		Average:	33.730	68.558	77.359

* True relative efficiency, e, ranges from 0.866 to 1.387.

TABLE III.VI.11

(β=2.0; σ^2=2.0; n=30)*

λ	Π	$\bar{\hat{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0100	0.423	5.730	5.885	6.086
0.50	0.0275	0.548	9.620	6.759	7.234
0.60	0.0500	0.623	3.800	6.772	3.440
0.70	0.1425	0.746	6.560	8.822	9.167
0.80	0.2850	0.867	8.360	10.103	6.584
0.90	0.3825	0.976	8.430	11.852	12.666
0.92	0.4050	0.987	7.240	12.075	12.598
0.94	0.4325	1.026	9.180	15.811	16.647
0.96	0.4600	1.031	7.380	13.460	14.157
0.98	0.4875	1.064	8.530	12.393	12.967
1.00	0.5575	1.120	12.000	20.048	21.744
1.02	0.5500	1.121	9.940	17.716	18.888
1.04	0.5975	1.132	8.850	16.578	17.427
1.06	0.5975	1.142	7.710	15.421	16.202
1.08	0.6500	1.180	9.310	14.244	15.338
1.10	0.6450	1.186	7.790	14.888	15.719
1.20	0.7675	1.280	6.670	14.325	14.908
1.30	0.8125	1.410	8.460	19.010	20.139
1.40	0.8575	1.492	6.580	19.498	20.112
1.50	0.9150	1.626	8.430	19.236	20.368
1.60	0.9225	1.686	5.360	12.967	13.104
		Average:	7.900	13.708	14.071

* True relative efficiency, e, ranges from 0.968 to 1.118.

TABLE III.VI.12

(β=3.0; σ^2=2.0; n=10)*

λ	Π	$\bar{\bar{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0700	0.518	29.550	33.973	37.477
0.50	0.1125	0.618	23.560	23.672	26.495
0.60	0.2275	0.774	28.920	37.190	42.629
0.70	0.2900	0.895	27.860	49.850	55.381
0.80	0.4425	1.111	38.880	74.878	87.191
0.90	0.5300	1.248	38.670	98.276	112.934
0.92	0.4500	1.183	28.620	69.321	78.023
0.94	0.5400	1.251	33.040	66.105	92.116
0.96	0.5450	1.333	38.850	63.013	72.635
0.98	0.5325	1.311	33.820	70.251	80.292
1.00	0.5375	1.280	27.960	93.656	101.626
1.02	0.5500	1.315	28.900	79.319	88.021
1.04	0.5975	1.392	33.830	84.467	96.504
1.06	0.5750	1.385	30.670	87.800	98.007
1.08	0.5675	1.398	29.440	85.142	94.881
1.10	0.6275	1.438	30.700	83.063	94.128
1.20	0.6700	1.491	24.280	73.795	81.244
1.30	0.6950	1.697	30.570	84.021	91.249
1.40	0.7950	1.950	39.270	91.322	113.251
1.50	0.8500	1.923	28.210	85.234	97.557
1.60	0.8925	2.133	33.330	109.918	128.745
		Average:	31.380	73.541	84.304

* True relative efficiency, e, ranges from 0.936 to 1.218.

TABLE III.VI.13

(β=3.0; σ^2=2.0; n=30)*

λ	Π	$\bar{\bar{\lambda}}$	% Bias($\hat{\lambda}$)	% Var.($\hat{\lambda}$)	% MSE($\hat{\lambda}$)
0.40	0.0050	0.430	7.530	5.013	5.244
0.50	0.0175	0.538	7.500	5.113	5.400
0.60	0.0650	0.652	8.750	8.269	8.770
0.70	0.1350	0.778	11.160	7.548	8.467
0.80	0.2250	0.862	7.700	10.099	10.594
0.90	0.3875	0.962	6.930	9.218	9.785
0.92	0.4425	1.007	9.430	11.066	11.941
0.94	0.4750	1.038	10.440	11.912	13.141
0.96	0.4700	1.027	7.010	10.217	10.781
0.98	0.5075	1.033	5.370	9.026	9.416
1.00	0.5425	1.067	6.720	9.536	10.184
1.02	0.5600	1.076	5.460	9.916	11.377
1.04	0.6200	1.124	8.130	10.274	11.015
1.06	0.6225	1.144	7.920	10.504	11.176
1.08	0.6325	1.148	6.270	9.584	10.138
1.10	0.6450	1.190	8.170	12.223	13.040
1.20	0.7750	1.260	5.040	10.102	10.461
1.30	0.8625	1.419	9.180	16.060	17.264
1.40	0.9100	1.525	8.960	15.598	16.917
1.50	0.9150	1.569	4.570	15.578	15.997
1.60	0.9700	1.692	5.740	16.046	16.821
		Average:	7.520	10.614	11.330

* True relative efficiency, e, ranges from 0.986 to 1.056.

TABLE III.VI.14

Estimated Power Functions*

Table	$\hat{\alpha}_o$	R^2
III.VI.2	0.7168	0.7540
	(0.0345)	
III.VI.3	1.1454	0.7294
	(0.1305)	
III.VI.4	0.7521	0.7906
	(0.0489)	
III.VI.5	1.3299	0.7565
	(0.1680)	
III.VI.6	0.7816	0.7800
	(0.0531)	
III.VI.7	1.4292	0.8766
	(0.1982)	
III.VI.8	0.7647	0.7921
	(0.0490)	
III.VI.9	1.0145	0.7111
	(0.0889)	
III.VI.10	0.7659	0.8014
	(0.0475)	
III.VI.11	1.1869	0.7858
	(0.1343)	
III.VI.12	0.7865	0.7703
	(0.0552)	
III.VI.13	1.2570	0.7575
	(0.1567)	

* Estimated standard errors appear in parentheses.
R^2 is calculated from residuals based on Π, not log Π.

"Accept $H_0 : \lambda \leq 1$ iff $\hat{\lambda} \leq 1$",

and the O.L.S. estimates of the power curves are summarized in Table III.VI.14. These estimates are based on the functional form

$$\Pi = \exp[-\alpha_0/\lambda],$$

or,

$$\log(\Pi) = -\alpha_0(1/\lambda).$$

The estimated power curves follow the anticipated general shape, but they suggest that the above test procedure is not very powerful in small samples. The power curves associated with Tables III.VI.2 and III.VI.3 appear in Figure III.VI.1 for illustrative[35] purposes.

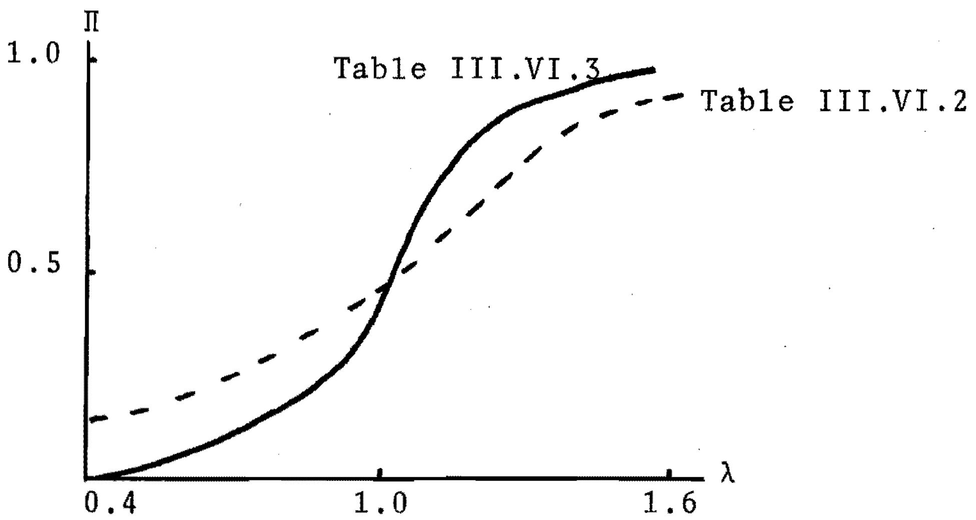


Figure III.VI.1: Estimated power curves.

In many of the situations studied in the experiment there is little difference between the true efficiencies of $\tilde{\beta}$ and $\hat{\beta}$, so the outcome of the $\hat{\lambda}$ (or $\tilde{\lambda}$) test may be of little practical consequence. However, in some of the situations

35. Of those cases where n = 10, the test based on $\hat{\lambda}$ in Table III.VI.2 is the least powerful. The corresponding results for n = 30 appear in Table III.VI.3.

tested there is up to 60% difference between the efficiencies of $\tilde{\beta}$ and $\hat{\beta}$, so the outcome of the $\hat{\lambda}$ test is more consequential.

(d) Summary

The above discussion concerning an ad hoc observable substitute for the unknown λ may be summarized as follows. First $\hat{\lambda}$ as defined in III.VI.3 is consistent but upward-biased by an amount exceeding $L = \text{tr.}\{[I + 2X'XA^{-1}]^{-1}\}$. This suggests that a superior substitute for λ would be $\tilde{\lambda} = (\hat{\lambda}-L)$, which is also consistent but has smaller bias than $\hat{\lambda}$ in small samples. In large samples $\tilde{\lambda}$ may be used with confidence, and increasingly so, since $L \to 0$ as $n \to \infty$.

Secondly, although $\tilde{\lambda}$ is upward-biased in small samples, it is still a useful statistic, since if a value of $\tilde{\lambda}$ substantially less than unity is computed in a particular problem, one may be fairly certain that in fact $\lambda < 1$. For values of $\tilde{\lambda}$ in excess of unity a somewhat greater margin would be needed before $\lambda > 1$ could be inferred with the same degree of confidence.

Finally, the results of the limited Monte Carlo experiment suggest that at least in some situations the small-sample biases and variances associated with $\hat{\lambda}$ or $\tilde{\lambda}$ may be quite substantial. These results should be treated cautiously as they refer only to specific situations and are limited to the one-regressor model.

VII. CONCLUSIONS

In this Chapter we have compared two M.E.L. estimators of the parameters in the multiple regression model, on the basis of a strong M.S.E. criterion, of which a commonly encountered weaker criterion is a special case. We have established a condition on the parameter space under which $\tilde{\beta}$ is "preferred" to $\hat{\beta}$, but this condition is asymmetric, a fact which has some interesting implications.

We have investigated (analytically) some of the effects of varying the parameters in the prior p.d.f. for β in the general model, and have shown some relationships between our analysis and that associated with the interpretation problem under collinear data.

We attempted to obtain a formal test of the hypothesis "$\tilde{\beta}$ is preferred to $\hat{\beta}$", but this was successful only for a special degenerate case. However, this case has some interesting relationships with other work by Toro-Vizcarrondo and Wallace and by Judge _et al_. (1973).

A simple _ad hoc_ substitute test statistic was proposed and was found to be consistent, but upward-biased in small samples. Since an observable lower bound for this bias was obtained, a second consistent, but still upward-biased, test statistic was suggested.

A small Monte Carlo experiment, limited to the one-regressor model, suggested that the magnitude of the bias could be substantial in small samples.

The M.S.E. comparison formulated in this Chapter provides part of the theoretical motivation in Chapter IV, where we consider the problem of seasonally adjusting an economic time-series.

CHAPTER IV

BAYESIAN SEASONAL ADJUSTMENT OF ECONOMIC TIME-SERIES

I. INTRODUCTION

This Chapter considers the problem of seasonally adjusting an economic time-series, the basis for discussion being the contribution by Jorgenson (1964). Thus, the presentation is in terms of a parametric decomposition of the time-series, involving simple dummy variables, rather than decomposition by linear filters. The adjustment problem is distinct from that of estimating an econometric relationship involving seasonal data, as is apparent from the debate between Lovell (1966) and Jorgenson (1967). The latter problem is mentioned only briefly in Section VII of this Chapter.

The Chapter argues for a Bayesian (or M.E.L.) analysis of the adjustment problem, for three reasons. First, the adjusted series may have more desirable statistical properties (*e.g.* in terms of M.S.E.) than if Jorgenson's classical analysis is applied, and the analysis in Chapter III provides some motivation here. Secondly, *a priori* information about the seasonal and non-seasonal components may be incorporated formally; and thirdly, certain model-comparison problems that are likely to arise may be treated more formally than is possible by classical methods.

The approach suggested here is used in a practical application in Chapter VIII, so that this Chapter partially bridges the two main parts of the thesis, as described in Section II of Chapter I.

II. EXPLAINING SEASONALITY

Let y be an n - vector of observations on the time-series to be seasonally adjusted. Under the traditional assumption of a linear additive[1] seasonal influence:

$$y = Da + Sb + \varepsilon \qquad \text{IV.II.1}$$

where:

D is an (n×q) matrix of observations such that rank (D)=q, and Da represents the systematic components[2] of y.

S is an (n×g) matrix of seasonal dummy variables, such that $S=(S_1, S_2, \ldots, S_g)$, where S_i is an n -vector of zero-one observations with $S_{it}=1$ iff period t is in season i. Thus rank (S)=g. Formulating S in this way assumes that the seasonality in y is <u>fixed</u> over the sample period. This may be inappropriate for a time-series covering a long period.

$\varepsilon_t \sim NI\ (0,\sigma^2)$; $t=1,2, \ldots, n$.

The "true" (unknown) seasonally adjusted series is taken as

---

1. Other assumptions are possible. For example, the "ratio-to-moving average" methods implicitly assume a multiplicative component. In such cases a logarithmic transformation leads to an additive model.
2. The distinction between the systematic and seasonal components of y may be somewhat subjective, and the decomposition of a time-series may be a difficult task. We abstract from such problems and treat S and D as well-defined matrices.

$$y_s = y - Sb, \qquad \text{IV.II.2}$$

which assumes[3] that the seasonality in y is of a strictly periodic nature. If $\tilde{b}$ is some estimator of b, then the corresponding estimator of y_s is

$$\tilde{y}_s = y - S\tilde{b} \qquad \text{IV.II.3}$$

Clearly, the choice of $\tilde{b}$ affects the form and the properties of $\tilde{y}_s$. Under the above assumptions Jorgenson proves[4] that

$$\hat{y}_s = y - S\hat{b}$$

is the unique B.L.U. estimator of y_s iff $\hat{b}$ is the O.L.S. estimator of b in IV.II.1.

We limit our attention to two problems: obtaining $\tilde{y}_s$ as in IV.II.3, and making inferences about the vectors a and b.

III. STATISTICAL PROPERTIES

As is argued by Tiao and Box (1973), for example, many strict Bayesians are not concerned with the sampling properties of their estimators. However, these estimators have sampling properties which are often considered desirable, as is emphasised in Chapter III. Under general

3. See Sims (1973), pp.28-37.
4. See Jorgenson (1964), pp.695-703.

regularity conditions they converge to M.L. estimators in large samples, in which case they are consistent. In the present problem, if a^* and b^* are Bayes estimates of a and b in IV.II.1, and if y_s^* is obtained from IV.II.3, then asymptotically $y_s^* \to y_s$, since $b^* \to b$. In particular, the consistency of b^* ensures that[5]

$$\underset{n\to\infty}{\text{plim}}\ (y_s^*) = \underset{n\to\infty}{\text{plim}}\ (y - S_b^*) = y_s$$

In finite samples Bayes estimators are admissible and minimize average risk, though this need not be true for M.E.L. estimators. A M.E.L. estimator based on an informative prior p.d.f. may be biased[6] and non-linear, but any linear combination of its elements may have smaller variance[7] (in all regions of the parameter space) than has any linear combination of the elements of $\hat{b}$, because of the information in the prior p.d.f.. Thus, it may be possible to obtain a b^* such that any linear combination of its elements has smaller M.S.E. than has any linear combination of the elements of $\hat{b}$, at least in some regions of the parameter space, as is discussed in detail in Chapter III. Then if one believes that the true parameters lie in such a region, one may prefer to use b^* in place of $\hat{b}$ when estimating the seasonally adjusted series.

5. We make use of Theorem A.13 in Appendix I.
6. In general, it will be biased if the mean of the prior p.d.f. differs from the true value of the parameter.
7. This is true for the natural-conjugate Bayes estimator discussed in Chapter III, for example. See Proposition III.III.1 in particular.

This M.S.E. property applies equally to estimators of y_s. For generality, let b^1 and b^2 be <u>any</u> estimators of b which give rise to y_s^1 and y_s^2, from IV.II.3. Then:

$$M.S.E.(y_s^i) = [\mathcal{E}(y_s^i - y_s)(y_s^i - y_s)']$$

$$= S[M.S.E.(b^i)]S' \quad ; \quad i=1,2. \qquad \text{IV.III.1}$$

and b^1 is preferred to b^2 on the basis of (matrix) M.S.E. iff

$$M.S.E.(b^2) - M.S.E.(b^1) = P_1,$$

where P_1 is a p.s.d. matrix of order g. However,

$$M.S.E.(y_s^2) - M.S.E.(y_s^1) = SP_1S' = P_2,$$

and P_2 is p.s.d. (of order n) iff P_1 is p.s.d., since S has full column rank $g \leq n$. Thus, y_s^1 is preferred to y_s^2 on the basis of matrix M.S.E. iff b^1 is preferred to b^2 on this basis. From the above discussion, it may be possible to obtain a y_s^* such that any linear combination of its elements has smaller M.S.E. than any linear combination of the elements of $\hat{y}_s$ (at least in some parts of the parameter space). As before, this may lead to a preference for y_s^* over $\hat{y}_s$.

IV. THE SYSTEMATIC COMPONENT

The vector Da explains the trend and cyclical movements of y. In a classical context, Jorgenson's $\hat{y}_s$ is obtainable whatever form D takes, provided that the latter has full column rank. Other estimators may not need even this rank condition to be satisfied.[8] However, the construction of D is of practical interest.

Often, it is suggested that D should take the form:

$$D = (T, T^2, \ldots, T^q) \qquad \text{IV.IV.1}$$

where T is an $(n\times 1)$ time-trend vector.[9] If D takes the form IV.IV.1 then its columns are likely to be collinear to some degree and the O.L.S. estimates of a and b in IV.II.1 may be rather imprecise. The results of Part (3) in Section IV of Chapter III again suggest that b^* (and hence y_s^*) may be helpful in this respect. Clearly, D may be constructed in other ways, but even if it takes the form IV.IV.1, there remains the problem of specifying q.

In a classical analysis a sequence of D matrices of the form IV.I.V1 might be constructed, with q being reduced by one at each step of the sequence. Each of these matrices could be tested in an O.L.S. regression of IV.II.1, and the usual t-test and F-test might be used to determine which D matrix should be retained. However, a shortcoming here is that such a procedure is subject to the well known problem of pre-testing bias.[10]

---

8. For example, the natural-conjugate Bayes estimator discussed in Chapter III is computable even if rank $(D)<q$.
9. That is, $T'=(1,2,3, \ldots, n)$. Other choices of base-value and/or increment for T might be considered. In this case the B.P.O. analysis suggested in Section VI can be extended to cover the additional models so formulated.
10. For example, see Bock et al. (1973), and the references cited therein.

Other *ad hoc* classical procedures might be considered. However, in keeping with the Bayesian viewpoint of this thesis, q might be treated as a discrete random variable, or used to define separate models to which the B.P.O. analysis described in Chapter II may be applied. These possibilities are covered in the next two Sections.

V. ESTIMATING THE SEASONAL INFLUENCE

It is easily shown that the problem of obtaining the Bayes estimator y_s^* amounts to the problem of obtaining the corresponding b^*. Under finite average risk, the Bayes procedure amounts to minimizing posterior expected loss.[11] In general, choosing y_s^* to minimize posterior expected loss is equivalent to choosing b^* on the same basis.

For example, under quadratic loss, the loss function associated with y_s^* is

$$L(y_s, y_s^*) = (y_s - y_s^*)'Q(y_s - y_s^*)$$

for p.d.s. Q of order n. But, from IV.II.2 and IV.II.3:

$$L(y_s, y_s^*) = (b^* - b)'(S'QS)(b^* - b),$$

and S'QS is also p.d.s. since S has full column rank. If b^* is the mean[12] of the marginal posterior p.d.f. for b,

11. See Section III of Chapter II, and the associated references.
12. See Section III of Chapter II. In the case of an absolute error loss function, b^* is chosen as the median of this p.d.f..

then both b^* and y_s^* minimize posterior expected loss. Thus, we may concentrate on the M.E.L. or Bayes estimation of b.

Consider three possible situations with respect to D.

First, let D be of a general (known) form, not necessarily that in IV.IV.1. A Bayesian approach to the estimation problem requires the specification of a prior p.d.f., $p(a,b,\sigma)$, for the parameters of IV.II.1. The likelihood function for the observations on y is well-defined once the distribution of ε is specified, and it may be written as $\ell(a,b,\sigma|y)$. Then, by Bayes' Theorem:

$$p(a,b,\sigma|y) \propto p(a,b,\sigma)\ell(a,b,\sigma|y) \qquad \text{IV.V.1}$$

Joint inferences about the elements of a and b are based on

$$p(a,b|y) = \int_0^\infty p(a,b,\sigma|y)d\sigma \qquad \text{IV.V.2}$$

and marginal inferences about b are based on

$$p(b|y) = \int p(a,b|y)da \qquad \text{IV.V.3}$$

Secondly, let D take the form IV.IV.1, with q fixed and known. Conditional on q, the prior p.d.f. for the parameters is $p(a,b,\sigma|q)$, and the likelihood function is $\ell(a,b,\sigma|y,q)$. Then,

$$p(a,b,\sigma|y,q) \propto p(a,b,\sigma|q)\ell(a,b,\sigma|y,q) \qquad \text{IV.V.4}$$

and,

$$p(b|y,q) = \int p(a,b,\sigma|y,q)da.d\sigma \qquad \text{IV.V.5}$$

Finally, let D be specified by IV.IV.1, with q an unknown discrete random variable distributed independently of the other parameters, with prior p.m.f. $p(q)$. Then[13]:

$$p(b|y) \propto \sum_q p(q)\int p(a,b,\sigma)\ell(a,b,\sigma|y,q)da.d\sigma \qquad \text{IV.V.6}$$

$$= \sum_q p(b|y,q)p(q|y) \qquad \text{IV.V.7}$$

where,

$$p(q|y) \propto p(q)\int \ell(a,b,\sigma|y,q)p(a,b,\sigma)da.db.d\sigma \qquad \text{IV.V.8}$$

is the posterior p.m.f. for q.

In each of the above three cases the marginal posterior p.d.f. for b, as in IV.V.3, IV.V.5, or IV.V.7, contains all of the prior and sample information about this parameter. As noted above, once a loss function is specified, the point estimate b^* emerges as some feature of this posterior p.d.f.. The seasonally adjusted series follows immediately as

$$y_s^* = y - Sb^* \qquad \text{IV.V.9}$$

The choice of the prior p.d.f., $p(a,b,\sigma)$, warrants special mention. Unless it is chosen so that the integrations

13. The likelihood is expressed as conditional on q, since unless q is specified the matrix D cannot be constructed, and then the likelihood function cannot be obtained.

necessary to obtain $p(b|y)$ or $p(b|y,q)$ can be performed analytically, computational burden is likely to be so great as to render this Bayesian approach impractical. One possibility is to adopt Jeffreys' diffuse[14] prior p.d.f.. Then, under a quadratic loss function (whatever the form of D), Jorgenson's result emerges as a special case, since then $b^* = \hat{b}$. However, this relies on obtaining b^* as a M.E.L. estimate, not as a strict Bayes estimate.

The natural-conjugate prior p.d.f. is also analytically tractable, and under the assumptions in Section II of this Chapter it is the Normal-Inverted Gamma distribution, as introduced in Section II of Chapter III. The attractiveness of this prior p.d.f. here is strengthened by the comments concerning M.S.E. in Section III and the results of Chapter III. Further, this prior p.d.f. is flexible enough to represent a variety of degrees of prior information.[15]

VI. MODEL COMPARISONS

The use of B.P.O. analysis[16] for model comparisons was suggested in Section IV of this Chapter. The matrix D may be constructed in several ways, each defining a different model of the general form IV.II.1. Let $\mathcal{M} = \{M\}$ be the (countable) model-space introduced in Section V of Chapter II. If prior odds, $\{p(M_i)/p(M_j)\}$, are assigned to any pair of models from $\mathcal{M}$, then the B.P.O. are:

---

14. See Section II of Chapter III.
15. See Raiffa and Schlaifer(1961;Ch.3).
16. See Section V of Chapter II for a discussion of the underlying probability and decision calculus.

$$\{p(M_i|y)/p(M_j|y)\} = \{p(M_i)/p(M_j)\}.\{p(y|M_i)/p(y|M_j)\}, \qquad \text{IV.VI.1}$$

where

$$p(y|M_k) = \int p(a_k,b_k,\sigma_k|M_k)\ell(a_k,b_k,\sigma_k|y,M_k)da_k.db_k.d\sigma_k \qquad \text{IV.VI.2}$$

and

$$M_k : y = D_k a_k + Sb_k + \varepsilon_k$$

$$\varepsilon_{tk} \sim NI\ (0,\sigma_k^2), \text{ for all } t; \quad k=i,j. \qquad \text{IV.VI.3}$$

The usual features of B.P.O. analysis apply. Although the posterior odds in IV.VI.1 are independent of the specification of $\mathcal{M}$, individual posterior probabilities depend on the other models in $\mathcal{M}$. Further, once a loss function is specified it may be used with the posterior odds (or probabilities) to select one model from $\mathcal{M}$, if so desired. Then b^* and y_s^* can be based on this "most probable" model. Preferably, all of the models should be retained and used to generate y_s^* from a weighted average of the b^*'s, the weights being the posterior probabilities of the models.

An interesting special case arises when each D_k is defined as in IV.IV.1, so that the models in IV.VI.3 are augmented by the specifications:

$$D_k = (T,\ T^2,\ \dots,\ T^k) \qquad \text{IV.VI.4}$$

for $k=i,j;\ \ i\neq j$.

Then $p(M_k) = p(q=k)$, and at least some proper prior information must be incorporated into the prior p.d.f.'s in IV.VI.2, since M_i and M_j contain different numbers of parameters.[17] In this case if one model is selected from $\mathcal{M}$, this amounts to choosing a value of q, say q^*, and in this case b^* is derived from $p(b|y,q^*)$ in IV.V.5 and is used to construct y_s^* as in IV.II.3. However, if all of the models are retained for generating y_s^*, then b^* is obtained from

$$p(b|y) = \sum_k p(b|y,M_k)p(M_k|y)$$

$$= \sum_q p(b|y,q)p(q|y) \qquad \text{IV.VI.5}$$

and the result is just that in IV.V.7.

Thus, one advantage of the M.E.L. (or Bayesian) analysis is that it yields exact finite-sample methods for model comparisons of the type likely to arise when seasonally adjusting data. Finally, as is noted in Section V of Chapter II, if the prior p.d.f.'s are proper then in large samples the B.P.O. are <u>approximately</u> equal to the usual likelihood ratio which arises with the "nested" models if the D_k's are defined as in IV.VI.4.

VII. REGRESSION RELATIONSHIPS WITH SEASONAL DATA

This Chapter concentrates on the problem of seasonally adjusting a single economic time-series. An associated (but

---

17. Strictly speaking, this may rule out diffuse prior p.d.f.'s as far as <u>exact</u> results are concerned in this part of the analysis. See Gaver and Geisel (1974), pp.66-72.

distinct) problem is that of estimating an econometric relationship involving seasonal data, as is relevant to Chapter VIII.

Jorgenson (1967) shows that the latter problem amounts to that of estimating the coefficients in one of a set of structural equations, and proposes Two Stage Least Squares estimation. Bayesian procedures could be adopted here, again facilitating the formal incorporation of *a priori* information. Bayesian counterparts of several classical structural form estimators have been developed.[18] However, the necessary analysis for Bayesian comparisons between multi-equation models is as yet in its infancy.[19] Thus, in Chapter VIII we abstract from the problem of simultaneity bias, and apply the methods developed in this Chapter directly to the raw data series.

Another common problem arises when raw data series are unavailable but seasonally adjusted series are supplied, perhaps several "adjusted" versions of one or more series being available. If these are for regressors in the relationship of interest, then this amounts to a special case of the "alternative data series" model-comparison problem, and B.P.O. may be used, as is shown by Geisel (1970), for example. However, B.P.O. analysis may not be meaningful if these different adjusted series are for the regressand in the relationship, since then it may not be sensible to compare

---

18. See Zellner (1971), Chapter 9, and the references cited therein.
19. See Gaver (1974).

the likelihood functions.[20]

VIII. CONCLUSIONS

This Chapter suggests a simple Bayesian (or M.E.L.) method of seasonally adjusting an economic time-series. This has the advantage of allowing the flexible and formal introduction of a priori information when estimating the components of the series. Secondly, a Bayesian analysis eases any model-comparison problems that may arise; and thirdly, the estimated seasonally adjusted series may have desirable statistical properties, these properties forming an alternative (and more general) basis to that proposed by Jorgenson. Finally, Jorgenson's result emerges as a special case of the M.E.L. approach to the adjustment problem.

In short, the simple Bayesian tools introduced in Chapter II are found to have yet another direct application to a common problem in econometric inference. This fact is exploited at a practical level in Chapter VIII, though some approximations are found to be necessary in that particular instance.

---

20. Meaningful comparisons are possible, for instance, if one regressand is a monotonic transformation of the other, as is discussed by Thornber (1966), pp.49-51. However, this is unlikely to be the case in the present problem.

CHAPTER V

BAYESIAN ANALYSIS OF DISTRIBUTED LAG MODELS: A SURVEY

I. INTRODUCTION

A distributed lag (D.L.) model is one in which the effect on the dependent variable of a change in an independent variable is spread over several (perhaps all) future periods. The model may contain several regressors, the impacts of which need not all be distributed over time.

For illustrative purposes, consider the simple D.L. model:

$$y_t = \sum_{0}^{\infty} \beta_i x_{t-i} + u_t \quad ; \quad t=1,2, \ldots, n. \qquad \text{V.I.1}$$

where the properties of u_t are specified by the analyst. Often it is convenient to make the (possibly strong) assumption that $\sum_{0}^{\infty} \beta_i$ is finite and that each β_i has the same sign, so that V.I.1. may be written:

$$y_t = \beta \sum_{0}^{\infty} w_i x_{t-i} + u_t \qquad \text{V.I.2}$$

with each $w_i \geq 0$, and $\sum_{0}^{\infty} w_i = 1$. Then $\{w_i\}$ may be interpreted as a set of probabilities, and the lag mean and variance may be obtained, if desired.

Lag operator notation may be adopted[1], so that V.I.2 becomes:

$$y_t = \beta W(\Lambda)x_t + u_t \qquad \text{V.I.3}$$

where

$$\Lambda^i(x_t) = x_{t-i}$$

and

$$W(\Lambda) = \sum_{o}^{\infty} w_i \Lambda^i \qquad \text{V.I.4}$$

The study of D.L. models was initiated by Fisher (1937), and subsequent contributions may be categorized according to their emphasis on economic theory, econometric theory, or applications. The first of these categories has been surveyed by Nerlove (1972); the second by Griliches (1967), Sims (1973), and Giles (1973); and all three by Wallis (1969).

One of the main difficulties with D.L. models is that the parameter space must somehow be condensed if estimation is to be feasible. Inevitably, this has been achieved in general by rather artificial means, as will be apparent from the discussion below.

D.L. models are common in economic studies and they give rise to a number of important problems for econometric inference. Some of these problems are discussed, from a frequentist viewpoint with the emphasis on asymptotic

1. See Dhrymes (1971a), Chapter 2.

results, by Dhrymes (1971a), but more recently a number of important Bayesian contributions have been made in this area. The purpose of this Chapter is to survey this Bayesian literature.

II. INFINITE LAG MODELS

(1) Rational Lags

Clearly, the D.L. model in V.I.2 or V.I.3 cannot be estimated in unrestricted form. Any constraints that are imposed should be based on prior knowledge, or else mis-specification biases may arise[2]. The B.P.O. analysis introduced in Chapter II is valuable in this context, and helps to overcome many of the conceptual problems faced by frequentists.

The class of constraints considered here is one in which the shape of the infinite lag distribution is restricted to some generating family which is specified by a relatively small number of parameters. The family that has received most attention in the literature is that of rational lag schemes, as proposed by Jorgenson (1966).

In this case, $W(\Lambda)$ from V.I.3 may be expressed as the ratio $A(\Lambda)/B(\Lambda)$, where:

$$A(\Lambda) = a_0 + a_1\Lambda + a_2\Lambda^2 + \dots + a_g\Lambda^g$$

$$B(\Lambda) = b_0 + b_1\Lambda + b_2\Lambda^2 + \dots + b_h\Lambda^h \qquad \text{V.II.1}$$

---

2. See Dhrymes, op.cit., Chapter 3.

Any arbitrary function may be approximated by a rational form.

If $W(\Lambda) = (1-\lambda)^r/(1-\lambda\Lambda)^r$, for $0 \leq \lambda < 1$ and $r \geq 1$, then this describes the Pascal family of D.L. models, proposed by Solow (1960). If $r = 1$, then the geometric D.L. model proposed by Koyck (1954) emerges as a special case of the Pascal (and hence of the general rational) class.

If $W(\Lambda)$ is assumed to be rational, then there are two ways of estimating V.I.3 - either directly, or after transformation to the autoregressive form:

$$B(\Lambda)y_t = \beta A(\Lambda)x_t + B(\Lambda)u_t \qquad \text{V.II.2}$$

The disturbance term in V.II.2 may exhibit a Markov process or a moving average process, depending on the assumptions made about u_t when V.I.1 is specified.

Estimation of the D.L. model is feasible now that $W(\Lambda)$ is constrained to a known family, provided that the number of specification parameters[3], whose values are unknown, is small. For particular values of these parameters, synthetic variables corresponding to the unknown initial-value parameters may be generated[4], and M.L. estimates may be computed in the manner described by Zellner and Geisel (1970),

3. For example, r and λ are the specifications parameters for the Pascal family.
4. The initial-value parameters are $\eta_j = \mathcal{E}(y_{-j})$, $j=1,2,..,h$; where h is the degree of $B(L)$. The synthetic variables z_{jt}, are $(h+1)$ in number, and are such that V.I.3 may be written:

$$y_t = \beta z_{0t} + \sum_1^h \eta_j z_{jt} + u_t$$

Rao (1971), and Maddala and Rao (1971). Such M.L. estimators may be justified on Bayesian grounds in large samples. Typically, this M.L. estimation involves a search over the specification parameter space. Dhrymes et al. (1970) provide an iterative search algorithm for the direct estimation of general rational D.L. models, this being based on earlier work by Steiglitz and McBride (1965) and Dhrymes (1969), (1971).

Bayesian methods are well suited for the incorporation of additional a priori information into the direct estimators. Such information may concern the initial-value parameters, for example, and Bayesian interpretations of this estimation problem are considered by Maddala (1971), Zellner and Geisel, and by Rao.

The Bayes estimator may be computed conditionally upon specific values of (some of) the specification parameters. Different values imply different models, thus giving rise to the problem of comparing and selecting among non-nested models. The B.P.O. analysis described in Chapter II has been found convenient in such cases and has been used in the context of simple[5] D.L. models by Thornber (1966), Chetty (1971), Geisel (1970), Zellner (1971), Lempers (1971), Maddala, and by Zellner and Geisel. The analysis is extended to the general rational class, and to allow for various error specifications, by Courville and Geisel (1971).

---

5. These authors consider nothing more general than the Pascal family of D.L. models. Zellner and Geisel consider alternative error specification in some detail.

(2) Other Contributions

Recently there have been several important contributions to the theory of estimating infinite D.L. models, not necessarily restricted to the class of rational models. The increasing attention being given to spectral techniques to analyse what is, after all, a time-series problem has been reflected in the Bayesian literature. A fundamental notion in spectral analysis is that of a "transfer function". Such functions[6] find frequent application in engineering, and recently they have been used to analyse D.L. models in economics. Transfer functions are directly analogous to rational D.L. models, and have been used in a Bayesian multi-equation environment by Zellner and Palm (1974), for example. Sims (1972a) makes use of spectral techniques to investigate seasonality in D.L. models, and places Bayesian interpretations on some of his results. Other contributions to the estimation of D.L. models by Sims involving some Bayesian element relate to estimation under prior restrictions (1972); discrete-time approximations in continuous-time models (1971a); and the difficulties imposed by the infinite-dimensional parameter space which arises with infinite D.L. models (1971). With regard to the latter, Sims is optimistic over the potential of Bayesian methods[7].

At least two features of rational D.L. models have led to a search for alternative approaches. The simpler

6. For example, see Jenkins and Watts (1968), p.47; Box and Jenkins (1971), Chapter II.
7. See Sims (1973), pp.40-45.

models in that family are very restricted in shape, while the more general ones involve many unknown parameters. Hall and Sutch (1968) develop a structure which has a flexible shape for short lags, yet allows for an infinite tail. Their estimator attempts to combine the best features of the rational family (for the tail) and the polynomial family[8] (for the head). The point at which the head and tail of the lag join must be chosen a priori, leading to a model-selection problem which could be handled by B.P.O. analysis, though this possibility is not suggested by those authors. Tsurumi (1973) suggests gamma, Poisson and generalized gamma distributions as alternatives to the rational family, claiming ease of estimation. Standard Bayesian methods of analysis are shown to apply easily to the resulting models.

III. FINITE LAG MODELS

An alternative way of restricting the model in V.I.2 or V.I.3 is to approximate the infinite lag series by finite truncation. That is, set $w_i = 0$ for all $i > L$, where L is finite, and where positive degrees of freedom are ensured by choosing $L < n$. Of course, a finite lag may be postulated initially, instead of V.I.1, rather than arise as an approximation to that infinite lag. In either case, there may be substantial problems over the a priori specification of L, and of the distributional shape. Typically, there is insufficient prior information to make

8. See Section III below for a discussion of the polynomial family.

these specifications straightforward, but again a Bayesian approach is helpful.

A particular exact specification suggested by Almon (1965), (1968) is to restrict the β_i's in V.I.1 to lie on a polynomial of specified degree, P. This proposal is discussed in detail in Chapter VII of this thesis. Given L and P, any further restrictions[9] that are imposed on the lag distribution may exactly determine its shape and modal position[10]. Once the value of L is chosen, P and any restrictions are selected so that the lag shape is consistent with *a priori* information.

The correct incorporation of such information is one of the practical difficulties with the classical version of this estimator, and the costs of mis-specification are substantial. Zellner and Williams (1973) illustrate how Bayesian methods may be applied to the "direct" version of the Almon estimator, and their analysis is generalized in Chapters VII and VIII of this thesis.

Typically, prior information is rather weak, so it may be difficult to choose L, P and any restrictions that are used. However, it should be possible to restrict the number of alternative specifications to a small set which is believed to include the true specification. This situation poses one of the greatest difficulties for the classical Almon procedure, since in that case it is meaningless to attribute probabilities to the alternative models,

---

9. Such restrictions are invariably linear, and often homogeneous. In practice they usually relate to the end-points of the lag distribution.
10. If P is altered, then typically the constraints also must be altered if the lag shape is to be unchanged. Fair and Jaffee (1971) provide a detailed analysis.

and pre-testing biases complicate the issue still further.

B.P.O. analysis is suggested in principle in this context by Maddala[11] and is alluded to by Zellner and Williams. The details are discussed in Chapter VII of this thesis, and some practical implications are shown in Chapter VIII.

Chetty provides a general Bayesian analysis of truncated versions of the Solow model estimated in direct form. Uniform prior densities are specified for r, λ and L, and alternative models are compared on the basis of various marginal and conditional posterior distributions. Chetty's approach is discussed[12] by Maddala and by Palm (1972).

Other contributions to the Bayesian analysis of finite D.L. models are made by Leamer (1970), (1972). One possibility suggested by Leamer is to use an <u>independent</u> Normal-Gamma prior distribution[13], with a proportional covariance matrix, but this approach makes heavy demands in terms of prior information. Maddala[14] discusses this suggestion, as well as that of Shiller (1973) that prior constraints be imposed on differences between the coefficients. Shiller's prior has a smoothing effect on the coefficients, but this is sometimes found to lead to nonsensical results. As prior uncertainty increases, Shiller's approach becomes equivalent to Almon's method.

---

11. Maddala, <u>op.cit.</u>, especially pp.12-13; pp.16-17.
12. Maddala, <u>op.cit.</u>, p.17; pp.21-22.
13. The usual natural-conjugate prior distribution is the product of a <u>conditional</u> Normal and an <u>independent</u> inverted gamma distribution.
14. Maddala, <u>op.cit.</u>, p.14.

Smoothing constraints also form the basis of the estimator proposed by Cleveland (1971), who considers the class of estimators lying between the M.L. estimator and an "extreme smoothness" estimator. Relative likelihood techniques are used to select a "partial smoothness" estimator from this class, and the functional form of this estimator is identical with that of a M.E.L. estimator based on a natural-conjugate prior distribution and quadratic loss function.

Palm[15] adopts Shiller's smoothing prior and uses it in conjunction with the "mixed" (continuous/discrete) prior analysis on a general finite D.L. model. He also proposes the use of the binomial distribution for the weights in a finite D.L. model, and provides the associated Bayesian analysis.

IV. CONCLUSIONS

Distributed lag models have received a lot of attention in the economic literature. The estimation of such models poses substantial problems for the frequentist, but in general these problems may be overcome to some extent in a formal and unified way if Bayesian methods are adopted.

Among the Bayesian contributions to D.L. analysis, most attention has focused on infinite lag structures. In part this reflects the historical development of the frequentist analysis of these models, but more importantly

15. Palm, op.cit., pp.8-13.

it reflects the advantages of the simplicity of the rational lag family. This simplicity is manifested in parameter spaces of low dimension, an important consideration from the Bayesian viewpoint.

The Bayesian analysis of finite D.L. models has been of a more fundamental nature, and has been dominated by a few excellent contributions, particularly those of Leamer and Shiller. Although this direct[16] attack on finite lag structures makes heavy demands in terms of _a priori_ information, it is potentially extremely powerful.

The less direct approach suggested by Almon and given a Bayesian interpretation by Zellner and Williams, provides a means of reducing the dimension of the parameter space to manageable proportions. Here, as with the rational D.L. models, the imposition of exact restrictions poses misspecification dangers for the frequentist. To some extent these are alleviated in the Bayesian analysis by assigning probabilities to the resulting models, and acting on the basis of posterior information. The advantages of the Bayesian approach to estimation and model comparisons are discussed in the context of the Almon estimator in Chapters VII and VIII of this thesis.

Having now surveyed the recent Bayesian contributions to the estimation of D.L. models, the remainder of the thesis is devoted to providing Bayesian analyses of some specific aspects of such models.

16. That is, the lag structure is not restricted to belong to some simple class described by only a few parameters.

CHAPTER VI

DISCRIMINATING BETWEEN AUTOREGRESSIVE FORMS: A MONTE CARLO COMPARISON OF BAYESIAN AND AD HOC METHODS[1]

I. INTRODUCTION

Griliches (1967) discusses an interesting discrimination problem arising with the estimation of certain rational D.L. models, and suggests an ad hoc rule of thumb for a common practical case. This Chapter compares Griliches' proposal with an alternative discrimination technique, B.P.O, as discussed in Chapter II and suggested in this particular context by Maddala (1971).

The formal B.P.O. analysis is expected to outperform Griliches' rule of thumb, at least in some parts of the parameter space. However, the results of this Chapter quantify these relative performances, and indicate the extent to which they are sensitive to parameter space location. They also demonstrate some of the dangers of applying ad hoc methods in econometrics.

If one adopts conventional lag operator notation, as in Chapter V, then a rational D.L. model is one of the form V.I.3, with $W(\Lambda)$ given by V.I.4 and V.II.1. Often it is convenient to estimate the rational D.L. model in its autoregressive form, V.II.2. The time scheme followed

1. A paper based on the contents of this Chapter has been accepted for publication by the Journal of Econometrics.

by the disturbances in V.II.2 depends[2] on the assumptions made about u_t in V.I.3. The problem considered by Griliches centres on the D.L. model in V.II.2 and on a simple regression model with serially correlated disturbances.

Let G(L) and H(L) be any rational lag generating functions, and let ε_t be a serially independent disturbance term. Then in general it is impossible to distinguish between

$$G(L)y_t = \varepsilon_t \qquad \text{VI.I.1}$$

and

$$y_t = H(L)\varepsilon_t \qquad \text{VI.I.2}$$

However, if exogenous variables are added to VI.I.1 and VI.I.2 then it <u>may</u> be possible to distinguish between these models, though this may be a difficult task.[3]

A simple example (involving an exogenous variable) likely to be met in practice involves a Koyck (geometric) D.L. model and a regression model with disturbances generated by a first-order Markov process.

The latter model is:

$$y_t = \gamma x_t + u_t \qquad \text{VI.I.3}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad ; \quad |\rho| < 1 \qquad \text{VI.I.4}$$

2. In particular, it may be possible to assume such a form of serial correlation for the u_t in V.I.3 that the disturbances in V.II.1 are serially independent. However, it may be difficult to justify such an assumption.
3. See Griliches, <u>op.cit.</u>, p.35.

and ε_t is NID $(0,\sigma_1^2)$; $t=1,2, \ldots, n$.

Thus,[4]

$$y_t = \gamma x_t + \rho y_{t-1} - \gamma\rho x_{t-1} + \varepsilon_t \qquad \text{VI.I.5}$$

The Koyck model is a special case of V.I.3, with $W(\Lambda) = (1-\lambda)/(1-\lambda\Lambda)$, where $0 \leq \lambda < 1$. Thus,

$$y_t = \beta x_t + \lambda y_{t-1} + (u_t' - \lambda u_{t-1}') \qquad \text{VI.I.6}$$

If the u_t' follow a first-order Markov process[5] with parameter λ, then

$$u_t' - \lambda u_{t-1}' = \varepsilon_t' \qquad \text{VI.I.7}$$

where ε_t' is NID $(0,\sigma_2^2)$, and VI.I.6 may be written:

$$y_t = \beta x_t + \lambda y_{t-1} + \varepsilon_t' \qquad \text{VI.I.8}$$

The practical motivation for Griliches' analysis is as follows. Suppose that in fact the model which generated the given sample is described by VI.I.3 and VI.I.4. If VI.I.8 is fitted to the sample by O.L.S. it is likely to

4. Note the restriction on the coefficient of x_{t-1} in VI.I.5.
5. Often, such a strong assumption cannot be justified. In its absence, the O.L.S. estimates of the coefficients in VI.I.6 are inconsistent, and this has motivated the development of alternative estimators. Griliches ignores the complications arising if VI.I.7 is <u>not</u> assumed, and concentrates entirely on VI.I.8.

explain the data rather well[6], so that one may mistakenly assume that the sample was generated by a D.L. model.

Griliches suggests that after VI.I.8 has been estimated, one should also estimate[7]

$$y_t = \alpha_1 x_t + \alpha_2 y_{t-1} + \alpha_3 x_{t-1} + e_t, \qquad \text{VI.I.9}$$

and if $\hat{\alpha}_3$ is significant and close to $-(\hat{\alpha}_1 . \hat{\alpha}_2)$, then one should infer that the sample was generated by VI.I.5 and not by VI.I.8.

One difficulty with this suggestion is that Griliches does not quantify the expression "close to", and one is left with an ad hoc rule of thumb. This is Griliches' intention, as is clear from the example he gives. There is another source of ambiguity. Griliches suggests that if his proposal is pursued and $\hat{\alpha}_3$ is significantly positive, this may mean that the sample comes from a generalised Koyck model[8]:

$$y_t = \beta x_t + \delta \sum_{o}^{\infty} \lambda^i x_{t-i} + u''_t \qquad \text{VI.I.10}$$

$$u''_t = \lambda u''_{t-1} + \varepsilon''_t, \qquad \text{VI.I.11}$$

or,

$$y_t = \beta x_t + \lambda y_{t-1} + (\delta - \beta\lambda) x_{t-1} + \varepsilon''_t \qquad \text{VI.I.12}$$

---

6. If time-series data are being used, and the lagged dependent variable appears as a regressor, then good explanation of the data may arise even from a mis-specified regression model.
7. O.L.S. is a consistent estimator under the assumption that ε_t is NID $(0, \sigma_3{}^2)$, for all t.
8. Note that VI.I.11 also involves a very special assumption about the disturbances.

Now, there is no a priori reason why $(\delta-\beta\lambda)$ should be positive, so for consistency Griliches should consider the possibility of VI.I.12 even if $\hat{\alpha}_3$ is negative in VI.I.9. He fails to mention this possibility, and concentrates on VI.I.5 and VI.I.8.

Maddala also limits his analysis to VI.I.5 and VI.I.8, and suggests B.P.O. as a formal and operational means of discrimination.[9] Maddala compares B.P.O. with Griliches' rule of thumb in one practical example with actual economic data, finding that both favour[10] the D.L. model, VI.I.8, in that single instance. Maddala suggests that it would be interesting to see the value of the posterior odds when $\hat{\alpha}_3$ is "quite close" to $-(\hat{\alpha}_1.\hat{\alpha}_2)$.

The frequency with which Griliches' procedure chooses the correct model should increase as the sample size increases, since O.L.S. applied to VI.I.9 is a consistent estimator, regardless of whether VI.I.5 or VI.I.8 is the true model. The behaviour of B.P.O. as the sample size increases is discussed at the end of Chapter II.

In this Chapter, a Monte Carlo study is used to compare the two methods of discriminating between VI.I.5 and VI.I.8. The limitation of the study to these two structures is discussed in Section II, and the details of the Monte Carlo experiment appear in Sections III and IV. The results are presented and analysed in Section V. Three major points are covered: First, B.P.O. are operational

9. We call here on the analysis presented in Section V of Chapter II.
10. In the case of Griliches' rule, $\hat{\alpha}_3$ is "quite different" from $-(\hat{\alpha}_1\hat{\alpha}_2)$. See Maddala, op.cit., p.29.

and have a formal basis, while Griliches' proposal is (deliberately) ad hoc and arbitrarily quantified; secondly, Maddala's request is answered, both when the true model is VI.I.5 and when it is VI.I.8; and thirdly, one is able to judge the relative merits of the two methods in different areas of the parameter space, and for various small-sample situations.

II. B.P.O. ANALYSIS

Since the situation to be analysed in this Chapter is of the type discussed in Section V of Chapter II, the possibility of other relevant models need not be considered, and we can join Griliches and Maddala in concentrating on only two.

From VI.I.5 and VI.I.8 these are:

$$M_1 : y_t = \gamma x_t + \rho y_{t-1} - \gamma\rho x_{t-1} + \varepsilon_t \qquad \text{VI.II.1}$$

$$M_2 : y_t = \beta x_t + \lambda y_{t-1} + \varepsilon_t' \qquad \text{VI.II.2}$$

where ε_t is NID $(0,\sigma_1^2)$ and ε_t' is NID $(0,\sigma_2^2)$, for all t.

From II.V.2, the B.P.O. may be written:

$$p(M_1|y,y_0,x)/p(M_2|y,y_0,x) = A/B$$

where:

$$A = p(M_1)\int p(y|\gamma,\rho,\sigma_1,y_0,x)p(\gamma,\rho,\sigma_1)d\gamma.d\rho.d\sigma_1 \qquad \text{VI.II.3}$$

$$B = p(M_2)\int p(y|\beta,\lambda,\sigma_2,y_0,x)p(\beta,\lambda,\sigma_2)d\beta.d\lambda.d\sigma_2 \qquad \text{VI.II.4}$$

Note the inclusion of initial conditions[11] in the likelihood functions in VI.II.3 and VI.II.4. Ideally, a full Bayesian analysis would test a variety of loss structures, prior p.d.f.'s, and p.m.f.'s, but for the present study the following assumptions are made:

(i) L is symmetric, so that $L(M_1,\bar{M}_2) = L(M_2,\bar{M}_1)$, and thus, from II.V.6, M_1 is preferred to M_2 iff $p(M_1|y)/p(M_2|y) > 1$.

(ii) $p(M_1) = p(M_2) = p(M)$, since there is no a priori information that either model should be favoured.

(iii) The parameters of each model are independently and uniformly distributed over their respective ranges. That is[12]:

$$
\begin{aligned}
p(\gamma,\rho,\sigma_1) &= p(\gamma).p(\rho).p(\sigma_1) & & \\
p(\gamma)d\gamma &= d\gamma &;\quad & 0 \leq \gamma \leq 1 \\
p(\rho)d\rho &\simeq 0.5d\rho &;\quad & -1 < \rho < 1 \\
p(\sigma_1)d\sigma_1 &\propto (d\sigma_1/\sigma_1) &;\quad & 0 < \sigma_1 < \infty
\end{aligned}
$$

and,

$$
\begin{aligned}
p(\beta,\lambda,\sigma_2) &= p(\beta).p(\lambda).p(\sigma_2) & & \\
p(\beta)d\beta &= 1.d\beta &;\quad & 0 \leq \beta \leq 1 \\
p(\lambda)d\lambda &\simeq 1.d\lambda &;\quad & 0 \leq \lambda < 1 \\
p(\sigma_2)d\sigma_2 &\propto (d\sigma_2/\sigma_2) &;\quad & 0 < \sigma_2 < \infty
\end{aligned}
$$

11. See Zellner and Tiao (1964), pp.763-764.
12. Note that the uniform prior p.d.f.'s on the $\log(\sigma_i)$ are improper, but that the other prior p.d.f.'s are proper.

Under these assumptions, and that of normality for ε_t and ε_t', integration with respect to σ_1 and σ_2 can be performed analytically, yielding:

$$A = 0.5p(M)\int\{\sum_1^n(y_t-\gamma x_t-\rho y_{t-1}+\gamma\rho x_{t-1})^2\}^{-n/2}d\gamma.d\rho \qquad \text{VI.II.5}$$

$$B = p(M)\int\{\sum_1^n(y_t-\beta x_t-\lambda y_{t-1})^2\}^{-n/2}d\beta.d\lambda \qquad \text{VI.II.6}$$

The ranges of integration in VI.II.5 and VI.II.6 are determined by the assumptions in (iii) above, and A and B are evaluated by bivariate numerical approximation[13]. If one of M_1 or M_2 is to be chosen, then by assumption (i) above, M_1 is chosen if $(A/B) > 1$, and M_2 is chosen if $(A/B) < 1$.

Of course, the approximate relationship II.V.5 between the B.P.O. and the L.R., L*, could be exploited here, provided that proper prior p.d.f.'s were assumed for the parameters. In this case, a likelihood ratio test (L.R.T.) would discriminate between M_1 and M_2, but critical values for this test are not tabulated.

Further, for M_1 and M_2:

$$L^* = (\hat{\sigma}_2^{\,2}/\hat{\sigma}_1^{\,2})^{n/2}$$

where:

$$\hat{\sigma}_1^{\,2} = (\tfrac{1}{n})\sum_1^n(y_t-\hat{\gamma}x_t-\hat{\rho}y_{t-1} + \hat{\gamma}\hat{\rho}x_{t-1})^2 \qquad \text{VI.I.7}$$

---

13. A computer program was written to generate the data for the Monte Carlo study, compute the B.P.O., and analyse the O.L.S. estimate of equation VI.I.9 according to Griliches' proposal. The bivariate integration routine is based on Simpson's rule, and is similar to that described by Zellner (1971), pp.409-414.

$$\hat{\sigma}_2^{\ 2} = (\tfrac{1}{n})\sum_1^n (y_t - \hat{\beta}x_t - \hat{\lambda}y_{t-1})^2 \qquad \text{VI.II.8}$$

are the M.L. estimates for $\sigma_1^{\ 2}$ and $\sigma_2^{\ 2}$. Although VI.II.8 can be obtained by O.L.S., the non-linearity of M_1 means that VI.II.7 has to be evaluated by an iterative search, perhaps similar to that proposed by Hildreth and Lu (1960). The computational cost of such a search could be comparable to that of performing the bivariate numerical integration needed to evaluate VI.II.5 and VI.II.6. Since the B.P.O. provide an operational test procedure which takes explicit account of the loss structure, we choose not to evaluate the L.R. values here. The possibility of a L.R.T. is also considered in Part (3) of Section III below.

The large-sample property noted at the end of Chapter II indicates that the B.P.O. will more frequently select the correct model as the sample size is increased. Of greater practical consequence is a comparison of B.P.O. analysis and Griliches' rule of thumb in small samples. The Monte Carlo study described in the next two Sections facilitates such a comparison.

III. THE MONTE CARLO STUDY

The Monte Carlo study falls into two parts, one in which M_1 is used to generate the data, and one in which the data are generated according to M_2. Each part is divided into six experiments, arising from two alternative sample sizes and three alternative sets of parameter values in each part of the study. Samples of size ten and thirty are investigated, and the same $\{x_t\}$ series and y_0 are used in all of the experiments.

(1) <u>M_1 is True</u>

In each of the six experiments in this part of the study the following steps are taken:

(i) One hundred $\{\varepsilon_t\}$ series are drawn randomly from a known distribution, and the corresponding one hundred $\{y_t\}$ series are generated according to M_1, given the parameters, $\{x_t\}$ and y_0.

(ii) In each of the one hundred replications, the B.P.O. are computed for M_1 vs. M_2.

(iii) Equation VI.I.9 is estimated by O.L.S. for each of the one hundred replications, and $\hat{\alpha}_3$ is compared with $-(\hat{\alpha}_1.\hat{\alpha}_2)$ in each of two ways.

These two methods of comparison reflect a desire to retain Griliches' informal interpretation of the "distance" between $\hat{\alpha}_3$ and $-(\hat{\alpha}_1.\hat{\alpha}_2)$, rather than to compare the B.P.O. with a formal non-Bayesian procedure. Griliches does not propose a formal test, and the example he gives clearly indicates his <u>ad hoc</u> use of this "distance".

In the first method of comparison, the "closeness" of $\hat{\alpha}_3$ to $-(\hat{\alpha}_1.\hat{\alpha}_2)$ is measured in terms of the percentage discrepancy between these two values, by computing the percentage: $G=|100(\hat{\alpha}_3+\hat{\alpha}_1.\hat{\alpha}_2)/\hat{\alpha}_3|$. Three variants of this method of comparison are used:

Test G1(β%): The test procedure as proposed by Griliches. The sign of $\hat{\alpha}_3$ must be the same as that for $-(\hat{\alpha}_1.\hat{\alpha}_2)$; $\hat{\alpha}_3$ must be significantly different from zero (by a one-tail t-test at the β% significance level); and G must be less than some specified percentage, if M_1 is to be chosen as the true model.

Test G2(β %): A modification of Griliches' proposal. No sign restriction is imposed on $\hat{\alpha}_3$. The two-tail t-test on $\hat{\alpha}_3$ is performed at the β% significance level.

Test G3: A further modification of Griliches' proposal. No attention is paid to the sign or to the t-test on $\hat{\alpha}_3$. Only the percentage value of G is used. Thus, Griliches' suggestion is interpreted rather liberally.

The second method of comparison is based on the test of the simple hypothesis H_o: "M_1 is true", against the simple alternative, H_a: "M_2 is true". If the sign of $\hat{\alpha}_3$ is not the same as that of $-(\hat{\alpha}_1.\hat{\alpha}_2)$, then H_a is accepted. Otherwise a conventional test of hypotheses proceeds. Under H_o, $t(\hat{\alpha}_3) = (\hat{\alpha}_3-\alpha_3)/(s.e.(\hat{\alpha}_3))$ is Student-t with mean $-(\alpha_1.\alpha_2)$, while under H_a it is Student-t with mean zero. Denote $(\hat{\alpha}_3/(s.e.(\hat{\alpha}_3))$ by $\tilde{\alpha}_3$, and for the purposes of applying the test, approximate $-(\alpha_1.\alpha_2)$ by $-(\hat{\alpha}_1.\hat{\alpha}_2)$.

Prior ignorance concerning the identity of the true model motivates the imposition[14] of $p(I) = p(II)$, so that H_o is accepted if either (i) $\hat{\alpha}_3$ and $-(\hat{\alpha}_1\hat{\alpha}_2)$ are both positive, and $\tilde{\alpha}_3 > -\frac{1}{2}(\hat{\alpha}_1\hat{\alpha}_2)$;

14. $p(I)$ and $p(II)$ are the probabilities of Type I and Type II errors respectively.

or, (ii) $\hat{\alpha}_3$ and $-(\hat{\alpha}_1\hat{\alpha}_2)$ are both negative, and
$\tilde{\alpha}_3 < -\tfrac{1}{2}(\hat{\alpha}_1\hat{\alpha}_2)$.

Otherwise, H_a is accepted. This interpretation of the basic test procedure is denoted Test G4 in this Chapter.

(2) M_2 is True

The six experiments in this part of the study follow the same steps as outlined above. However, in this case the $\{y_t\}$ series are generated according to M_2.

(3) Limitations of the Tests

As noted already, the above tests are informal by design, since an objective of this Chapter is to test Griliches' *ad hoc* procedure. However, Griliches' problem could be altered slightly so that the two models compared were VI.I.5 and VI.I.12. The former is "nested" in the latter, and is obtained by restricting $\delta=0$ in VI.I.12. If this situation were analysed, a formal classical test is available.

A B.P.O. analysis of *these* two models could be made, and again these new B.P.O. are related to a L.R., as in Section II. The computational inconvenience of the non-linearity of VI.I.5, as noted in Section II, again applies when computing the L.R. for VI.I.12 and VI.I.5:

$$L^{**} = (\hat{\sigma}_1^{\,2}/\hat{\sigma}_3^{\,2})^{n/2}$$

where $\hat{\sigma}_1^{\,2}$ and $\hat{\sigma}_3^{\,2}$ are again M.L. estimates. However, in this case L^{**} can be used to discriminate between VI.I.5 and

VI.I.12. As is shown by Silvey (1970), 2 log L** is approximately distributed as $\chi_1^{\,2}$ in large samples, since one restriction is placed on VI.I.12 to obtain VI.I.5.

However, we are concerned primarily with VI.I.5 and VI.I.8 (not VI.I.12), as are Griliches and Maddala. Further, although the outcomes of the L.R.T. based on L** would be interesting, it should be no more expensive to compute the corresponding B.P.O. and thus remove the need for large-sample approximation.

The two methods outlined in Part (1) of Section III for comparing $\hat{\alpha}_3$ with $-(\hat{\alpha}_1\hat{\alpha}_2)$ are each somewhat inadequate, even though they are simpler to apply than the (formally correct) B.P.O. analysis.

The performance of either method improves as the magnitude of $|-(\alpha_1\alpha_2)|$ increases, and hence (at least for large samples) the magnitude of $|-(\hat{\alpha}_1\hat{\alpha}_2)|$ increases. As this value departs from zero, the first method is such that pr.$\{G<\beta\%\}$ increases, and the power of the test improves. Similarly, as the magnitude of $|-(\alpha_1\alpha_2)|$ increases, the second method is such that either pr.$\{\tilde{\alpha}_3>-\frac{1}{2}(\hat{\alpha}_1\hat{\alpha}_2)\}$ increases (if both $\hat{\alpha}_3$ and $-(\hat{\alpha}_1\hat{\alpha}_2)$ are positive), or pr.$\{\tilde{\alpha}_3<-\frac{1}{2}(\hat{\alpha}_1\hat{\alpha}_2)\}$ increases (if both $\hat{\alpha}_3$ and $-(\hat{\alpha}_1\hat{\alpha}_2)$ are negative). Again, the power of the test depends on the magnitude of $|-(\alpha_1\alpha_2)|$.

However, the two methods used are certainly preferable to the even more simplistic approach of judging between M_1 and M_2 solely on the basis of $|\hat{\alpha}_3+\hat{\alpha}_1\hat{\alpha}_2|$. The latter quantity measures the absolute deviation between the parameter estimates, but is always relative to the magnitudes of the true parameters. Any test based on $|\hat{\alpha}_3+\hat{\alpha}_1\hat{\alpha}_2|$ could be quantified (and hence applied) only if the analyst had some idea of the size of the true parameters.

IV. THE CHOICE OF DATA AND PARAMETERS

(1) Exogenous Variable

The exogenous variable series $\{x_t\}$, comprises a growth component and a random component. The former has a base value of 10.00 and grows by 2% per observation. The latter is drawn from the standard normal distribution.[15]

(2) Parameter Values

To ensure parameter values of realistic magnitudes, M_1 and M_2 are treated as models of consumption expenditure, and the parameter values are chosen to reflect a long-run marginal propensity to consume of 0.90.

In Part (1) of the study the three alternative parameter specifications are:

(i) $\gamma = 0.90$; $\rho = 0.30$

(ii) $\gamma = 0.90$; $\rho = 0.60$

(iii) $\gamma = 0.90$; $\rho = 0.90$

The values of ρ reflect the fact that positively serially correlated disturbances are very common in time-series analysis, and the value of γ satisfies the desired restriction on the long-run m.p.c..

15. All standard normal deviates used in this Chapter are from Rand Corporation (1950). The $\{x_t\}$ series is that used in Chapter III.

In Part (2) of the study the three alternative parameter specifications are:

(i) $\beta = 0.09$; $\lambda = 0.90$

(ii) $\beta = 0.45$; $\lambda = 0.50$

(iii) $\beta = 0.81$; $\lambda = 0.10$

These values satisfy the conditions that $\beta/(1-\lambda)=0.90$, and that $0 \leq \beta \leq 1$, $0 \leq \lambda < 1$, while testing a broad range of values for each parameter.

(3) Random Disturbances

The series $\{\varepsilon_t\}$ and $\{\varepsilon_t'\}$ are drawn from the standard normal distribution. This reduces the computational burden and implies "Hybrid R^2" values of realistic magnitudes. Since var.(ε) = var. (ε') = 1, the population R^2 is given by R^2 = [var.(y) - 1]/var.(y). In the case of M_1 this reduces to:

$$R_1^2 = [\gamma^2(1-\rho^2)\text{var.}(x) + \rho^2]/[\gamma^2(1-\rho^2)\text{var.}(x) + 1]$$

and in the case of M_2:

$$R_2^2 = [\beta^2\text{var.}(x) + \lambda^2]/[\beta^2\text{var.}(x) + 1]$$

Hybrid R^2 values may be calculated by substituting the true parameter values and the sample variance of $\{x_t\}$ into the formulae for R_1^2 and R_2^2. The results appear in Table VI.IV.1.

TABLE VI.IV.1

Hybrid R^2 values

Parameter Set	R_1^2		R_2^2	
	10 obs.	30 obs.	10 obs.	30.obs.
(i)	0.52	0.83	0.81	0.82
(ii)	0.61	0.84	0.40	0.66
(iii)	0.84	0.90	0.45	0.65

(4) Initial Value of y

Since the models M_1 and M_2 each involve the lagged dependent variable, an initial value, y_0, must be supplied[16] if $\{y_t\}$ is to be generated from $\{x_t\}$, the disturbances, and the given parameters. Taking 90% of the initial value of $\{x_t\}$ and adding a standard normally distributed random factor yields the value $y_0 = 8.12$.

V. RESULTS

The two parts of the study each comprise six experiments each of which is based on one hundred replications. The results are summarised in six tables, one for each of the different sets of true parameter values in conjunction with both alternative sample sizes. Within each experiment the four variations of Griliches' method of discrimination are compared with B.P.O. analysis.

16. See Zellner and Tiao, op.cit., for alternative assumptions which lead to the same a posteriori results.

The notation used in the tables below is:

N_B = number of times that the B.P.O. favour M_1.

N_g = number of times that the Test G4 favours M_1.

$N_G(\tau)$ = number of times that $G \leq \tau$, τ=5,10,25; so that M_1 is favoured by the tests G1 to G3.

Rho = Spearman's Rho statistic[17], measuring the rank correlation between the values of the B.P.O. and the values of $|\hat{\alpha}_3 + \hat{\alpha}_1.\hat{\alpha}_2|$ over the full one hundred replications.

Z = standard normal transformation of Rho.[18]

In computing Spearman's Rho statistic, the B.P.O. are ranked in ascending order and the values of $|\hat{\alpha}_3+\hat{\alpha}_1\hat{\alpha}_2|$ in descending order, so that in either case the higher the ranking the more M_1 is favoured over M_2.

(1) <u>M_1 is True</u>

Tables VI.V.1 to VI.V.3 show the results of the six experiments in the first part of the Monte Carlo study. In each case, the same value of γ is used, but ρ varies from 0.3 to 0.9. If the tabulated figures for N_G, N_g and N_B are subtracted from one hundred and then divided by one hundred, the resulting figures may be interpreted as probabilities of a Type I error. Thus, the <u>larger</u> the values of N_G, N_g and N_B, the more successful are the associated methods of discrimination.

17. See Mills (1955), pp.311-312.
18. Z is to be used to test the hypothesis that there is zero rank correlation between the values of the B.P.O. and of $|\hat{\alpha}_3+\hat{\alpha}_1.\hat{\alpha}_2|$. See Mills, <u>op.cit</u>., pp.315-316.

As is expected from the discussion in Part (3) of Section III above, the G-tests are all sensitive to the location of model M_1 in the parameter space. In general, each G-test performs better as the value of ρ (and hence of $\gamma\rho = \alpha_1\alpha_2$) increases. The consistency of O.L.S. is reflected in the fact that the G-tests perform better (for a given set of parameter values) as the sample size is increased.

The values for N_B are far more stable than those for N_G and N_g as the parameter values are changed, but for a given parameter set the B.P.O. are also more successful as the sample size is increased, reflecting the large-sample property noted at the end of Chapter II.

The range of $p(I)$ for B.P.O. analysis is from 0.00 to 0.38; for G4 it is from 0.00 to 0.69; and for the most stringent form of G1 it is from 0.15 to 1.00. Unless a high level of tolerance (a large value of G) is permitted with the tests G1 to G3, they tend to be strongly in favour of M_2, thus admitting a high $p(I)$.

For samples of size ten, the test as proposed by Griliches, G1, is never as successful as is B.P.O. analysis in this part of the study, though for samples of size thirty and $\rho \geq 0.60$, G1 can be as successful as are the B.P.O. in selecting the true model, but only if the very large tolerance level of 25% is permitted with G1.

The test G4 is uniformly more successful than are G1 to G3, and is more successful than is B.P.O. analysis in two of the six cases shown in Tables VI.V.1 to VI.V.3.

<u>TABLE VI.V.1</u>

M_1 True; γ=0.90; ρ=0.30

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	0	1	3	-			
	G1(10%)	2	4	6	-			
	G2(5%)	0	1	3	-	62		
	G2(10%)	0	1	3	-			
	G3	2	5	16	-		-0.107	-1.063
	G4	-	-	-	31			
n=30	G1(5%)	9	19	38	-			
	G1(10%)	11	22	42	-			
	G2(5%)	6	13	27	-	80		
	G2(10%)	9	19	38	-			
	G3	12	24	47	-		0.013	0.133
	G4	-	-	-	72			

TABLE VI.V.2

M_1 True; γ=0.90; ρ=0.60

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	4	11	26	-			
	G1(10%)	8	19	40	-			
	G2(5%)	3	8	15	-	83		
	G2(10%)	4	11	26	-			
	G3	9	20	56	-		0.159	1.583
	G4	-	-	-	71			
n=30	G1(5%)	41	64	98	-			
	G1(10%)	41	64	98	-			
	G2(5%)	40	63	97	-	97		
	G2(10%)	41	64	98	-			
	G3	41	64	98	-		0.051	0.503
	G4	-	-	-	100			

TABLE VI.V.3

M_1 True; γ=0.90; ρ=0.90

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	21	39	52	-			
	G1(10%)	22	43	68	-			
	G2(5%)	17	30	38	-	80		
	G2(10%)	21	39	52	-			
	G3	25	51	82	-		0.428	4.258
	G4	-	-	-	89			
n=30	G1(5%)	85	96	100	-			
	G1(10%)	85	96	100	-			
	G2(5%)	85	96	100	-	100		
	G2(10%)	85	96	100	-			
	G3	85	96	100	-		0.249	2.456
	G4	-	-		100			

However, overall the B.P.O. fare considerably better than does Griliches' proposal, unless one is prepared to interpret and apply the latter in a rather loose manner.

Maddala's request regarding the magnitude of the B.P.O. when $\hat{\alpha}_3$ is "close to" $-(\hat{\alpha}_1.\hat{\alpha}_2)$ is answered in Table VI.V.4. Shown there are the average values of the B.P.O. for those replications in which the strictest form of Griliches' test, G1(5%), is applied and $G \leq 5\%$ is satisfied.

TABLE VI.V.4

$G \leq 5\%$

Table	VI.V.1		VI.V.2		VI.V.3	
n	10	30	10	30	10	30
$N_G(5)$	0	9	4	41	21	85
$\overline{\text{B.P.O.}}$	-	1.39 $\times 10^1$	4.89 $\times 10^2$	2.76 $\times 10^4$	1.02 $\times 10^2$	4.88 $\times 10^6$

The B.P.O. favour the true model, M_1, whenever the strict G-test is applied at the 5% tolerance level, often discriminating to a remarkable degree. Of course, as is shown in the earlier tables, the B.P.O. favour the true model in a large number of replications when none of the G-tests perform particularly well.

Considering the rank correlation between the B.P.O. and $|\hat{\alpha}_3 + \hat{\alpha}_1.\hat{\alpha}_2|$, the null hypothesis, H_o': "zero rank correlation", cannot be rejected (at the 5% or 1% significance levels) in Tables VI.V.1 or VI.V.2, but in Table

VI.V.3 it must be rejected (at the 1% significance level) in favour of H_a': "positive rank correlation". In the latter case, the B.P.O. favour M_1 more strongly as the distance $|\hat{\alpha}_3 + \hat{\alpha}_1.\hat{\alpha}_2|$ decreases.

(2) M_2 is True

The results of the six experiments in the second part of the Monte Carlo study appear in Tables VI.V.5 to VI.V.7. In this case, the smaller the values of N_G, N_g and N_B, the more successful are the associated methods of discrimination, since if the tabulated figures are divided by one hundred they may be interpreted as probabilities of a Type II error.

Again, as expected from the discussion in Part (3) of Section III, the G-tests are all sensitive to the location of M_2 in the parameter space. In Tables VI.V.5 and VI.V.7, $\beta\lambda = 0.081$, while in Table VI.V.6 $\beta\lambda = 0.225$. Thus, it is not surprising to see smaller values for N_G and N_g in Table VI.V.6 than in Tables VI.V.5 or VI.V.7, while these values in Tables VI.V.5 and VI.V.7 are of similar magnitudes. Again, the consistency of O.L.S. is reflected in the improved performances of the G-tests (for a given set of parameter values) as the sample size is increased.

The values for N_B are again quite uniform (for a given sample size) as the parameter values are varied. Again, as expected, B.P.O. analysis performs better (for a given set of parameter values) as the sample size is increased.

The range of p(II) for B.P.O. analysis is from 0.05

to 0.32; for G4 it is from 0.13 to 0.29; and for the most stringent form of G1 it is from 0.00 to 0.02.

Even if a high level of tolerance is permitted with the tests G1 to G3, they strongly favour M_2, as they did in the first part of the study. As a result, these tests perform markedly better than does B.P.O. analysis, but the latter is superior to the test G4 in all but one instance.

Considering the rank correlation between the B.P.O. and $|\hat{\alpha}_3 + \hat{\alpha}_1.\hat{\alpha}_2|$, the null hypothesis, H_o': "zero rank correlation", is rejected (at the 1% significance level) in favour of H_a': "positive rank correlation", in all but one instance in Tables VI.V.5 to VI.V.7. The exception is in Table VI.V.7 (n=10), where H_o' cannot be rejected at the 5% significance level. In all other cases there is a high positive correlation between the magnitudes of the B.P.O. and the magnitudes of $|\hat{\alpha}_3 + \hat{\alpha}_1.\hat{\alpha}_2|$.

B.P.O. analysis is less successful both in absolute terms and relative to the tests G1 to G3 when M_2 is the true model, compared with when the true model is M_1. However, much of the apparent success of these G-tests results from their pronounced tendency to reject M_1 under any circumstances.

Since typically one is quite ignorant[19] about the identity of the true model in practical applications, this property of the tests G1 to G3 diminishes their usefulness, particularly when compared with B.P.O. analysis in samples of size thirty or more.

19. The analysis in this Chapter is based on such an assumption. If there is any a priori information in favour of either M_1 or M_2, then this should be incorporated in the specification of $p(M_1)$ and $p(M_2)$.

TABLE VI.V.5

M_2 True; β=0.09; λ=0.90

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	2	3	5	-	22	0.577	5.741
	G1(10%)	4	5	7	-			
	G2(5%)	1	2	3	-			
	G2(10%)	2	3	5	-			
	G3	5	8	15	-			
	G4	-	-	-	33			
n=30	G1(5%)	0	1	4	-	12	0.896	8.920
	G1(10%)	0	1	5	-			
	G2(5%)	0	0	1	-			
	G2(10%)	0	1	4	-			
	G3	0	3	7	-			
	G4	-	-	-	29			

TABLE VI.V.6

M_2 True; β=0.45; λ=0.50

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	0	1	2	-			
	G1(10%)	1	3	4	-			
	G2(5%)	0	1	1	-	32		
	G2(10%)	0	1	2	-			
	G3	2	8	12	-		0.757	7.528
	G4	-	-	-	28			
n=30	G1(5%)	0	0	1	-			
	G1(10%)	0	0	1	-			
	G2(5%)	0	0	1	-	5		
	G2(10%)	0	0	1	-			
	G3	0	0	1	-			
	G4	-	-	-	13		0.726	7.226

TABLE VI.V.7

M_2 True; β=0.81; λ=0.10

	Test	$N_G(5)$	$N_G(10)$	$N_G(25)$	N_g	N_B	Rho	Z
n=10	G1(5%)	2	2	4	-			
	G1(10%)	3	6	9	-			
	G2(5%)	2	2	3	-	24		
	G2(10%)	2	2	4	-			
	G3	3	8	14	-		-0.068	-0.679
	G4	-	-	-	29			
n=30	G1(5%)	0	1	3	-			
	G1(10%)	0	1	6	-			
	G2(5%)	0	0	2	-	11		
	G2(10%)	0	1	3	-			
	G3	0	4	8	-		0.448	4.458
	G4	-	-	-	26			

(3) Summary

The results in Tables VI.V.1 to VI.V.3 and Tables VI.V.5 to VI.V.7 are summarized in Table VI.V.8. The figures in the first three tables are converted to probabilities of a Type I error, $p_i(\mathrm{I})$, i=1,2,3. (The i-subscript denotes a figure in the ith. of the first three tables.) Similarly, the figures in Tables VI.V.5 to VI.V.7 are converted to probabilities of a Type II error, $p_j(\mathrm{II})$, j=1,2,3. (The j-subscript denotes a figure in the jth. of Tables VI.V.5 to VI.V.7.)

The cross-table averages $\overline{p(\mathrm{I})} = \frac{1}{3}\sum_i p_i(\mathrm{I})$, and $\overline{p(\mathrm{II})} = \frac{1}{3}\sum_j p_j(\mathrm{II})$ give a good indication of the errors likely to be incurred by applying each of the tests when either M_1 or M_2, respectively, is true.

Since the true model is unknown, corresponding values of $\overline{p(\mathrm{I})}$ and $\overline{p(\mathrm{II})}$ can be averaged to give the expected probability of committing an error (either Type I or Type II) when each of the tests is applied. These final averages appear in Table VI.V.8.

On this composite basis, B.P.O. analysis always performs markedly better than do any of the G-tests, the best of the latter generally being the test G4. In fact, the performances of the more stringent G-tests, with samples of size ten, are little better than could be attained by tossing a fair coin.

TABLE VI.V.8

$\frac{1}{2}[\overline{p(\mathrm{I})} + \overline{p(\mathrm{II})}]$

	Test	G < 5%	G < 10%	G < 25%	g	B
n=10	G1(5%)	0.465	0.425	0.383	-	0.225
	G1(10%)	0.443	0.413	0.343	-	
	G2(5%)	0.471	0.443	0.418	-	
	G2(10%)	0.465	0.425	0.383	-	
	G3	0.456	0.413	0.311	-	
	G4	-	-	-	0.331	
n=30	G1(5%)	0.275	0.205	0.120	-	0.085
	G1(10%)	0.271	0.201	0.121	-	
	G2(5%)	0.281	0.213	0.133	-	
	G2(10%)	0.275	0.205	0.120	-	
	G3	0.270	0.205	0.118	-	
	G4	-	-	-	0.160	

VI. CONCLUSIONS

B.P.O. analysis is a formal quantified method of model discrimination, for all sample sizes, and may be applied directly to the problem analysed in this Chapter. In comparison, Griliches' proposal is intentionally <u>ad hoc</u> and must be quantified informally, but once quantified it is easily applied in practice. In fairness to its proponent, it should be noted that the latter procedure was intended only as an easily applied rule-of-thumb and not as a formal test. As was shown in Part (3) of Section III, Griliches' problem can be altered so that a classical test is applicable.

The results and conclusions of Section 5 must be judged in the light of the assumptions on which the Monte Carlo study is based, and the particular choice of parameters, <u>etc</u>. The study could be extended by investigating alternative prior p.d.f.'s and prior p.m.f.'s and alternative error variances, for example, but the results obtained here suggest that the type of discriminatory procedure suggested by Griliches has a number of important shortcomings. These alone limit its usefulness in small-sample applications.

In particular, this type of <u>ad hoc</u> procedure tends to reject the serial correlation model in favour of the D.L. model, when "closeness" between $\hat{\alpha}_3$ and $-(\hat{\alpha}_1\hat{\alpha}_2)$ is measured in terms of a small percentage discrepancy. In this form, the "test" exhibits high p(I) and low p(II), both decreasing as the sample size is increased.

If the hypothesis-testing variant, G4, is adopted, the results are very similar, but a little inferior, to those obtained by B.P.O. analysis. In particular, the test G4 does not have the same strong tendency to favour the D.L. model that the other G-tests have.

If one has no prior information to favour either of the models, then it is risky to use G1, G2 or G3, especially when B.P.O. analysis is relatively straightforward in this case. If M_1 is favoured a priori, then B.P.O. analysis incorporating $p(M_1) > p(M_2)$ should be used. If M_2 is favoured a priori, then the G-tests described in this Chapter may offer some guidance, but it would be preferable to use B.P.O. analysis[20] incorporating $p(M_2) > p(M_1)$.

It would be interesting to see a comparison of B.P.O. analysis and a formal classical procedure applied to the generalized Griliches' problem noted in Part (3) of Section III, but this is not an objective of this Chapter.

Although the technique originally proposed by Griliches has been interpreted and modified rather liberally here, it is fair to say that this type of discriminatory procedure should be treated with caution, as should other ad hoc methods in econometrics.

---

20. Putting $p(M_2) > p(M_1)$ in this case will improve the performance of B.P.O. analysis in comparison with the results shown in Part (2) of Section V. In any case, one has the comfort of discriminating between M_1 and M_2 on a formal basis.

CHAPTER VII

BAYESIAN INFERENCE AND THE RESTRICTED ALMON ESTIMATOR. PART I : THEORETICAL RESULTS

I. INTRODUCTION

This Chapter is concerned with a general Bayesian interpretation of the classical Polynomial Approximation Estimator (P.A.E.), often known as the Almon estimator[1], for the general finite distributed lag model (F.D.L.M.).

The model involves r D.L. regressors and q other regressors, the latter possibly including a constant and/or dummy variables:

$$y_t = \sum_{j=1}^{r} \left(\sum_{k=0}^{L_j} w_{jk} x_{jt-k} \right) + \sum_{i=1}^{q} \phi_i c_{it} + \varepsilon_t \qquad \text{VII.I.1}$$

There are n observations on the $\{y_t\}$, $\{x_{jt}\}$ and $\{c_{it}\}$ series after lags are allowed for, and the ε_t are assumed to be NID$(0,\sigma^2)$ for all t, unless otherwise stated.

To simplify the exposition, we often refer to a special case of VII.I.1:

$$y_t = \sum_{k=0}^{L} w_k x_{t-k} + \varepsilon_t \qquad \text{VII.I.2}$$

The recent Bayesian contribution by Zellner and Williams (1973), hereafter ZW, forms the starting point for

---

1. See Almon (1965), and Cooper (1972), for example.

this Chapter. The limitations of the classical P.A.E. and the ZW approach are analysed, and the latter procedure is generalized in several ways.

II. THE CLASSICAL P.A.E.

Consider model VII.I.2, but note that the discussion in this Section and the next applies equally to each of the r D.L. variables in VII.I.1.

(1) Estimation

In the classical P.A.E. the discrete distribution of the w_k's is replaced by a continuous function, and this function is approximated[2] by a polynomial of degree P in the interval [0,L]. Thus,

$$f(k) \simeq \sum_{j=0}^{P} \alpha_j k^j \quad ; \quad k=0,1,\ \ldots,\ L. \quad P\leq L \qquad \text{VII.II.1}$$

Disregarding approximations, VII.I.2 becomes:

$$y_t = \sum_{j=0}^{P} \alpha_j \left(\sum_{k=0}^{L} k^j x_{t-k} \right) + \varepsilon_t \quad ,$$

or,

$$y_t = \sum_{j=0}^{P} \alpha_j \psi_{jt} + \varepsilon_t \quad , \qquad \text{VII.II.2}$$

where:

$$\psi_{jt} = \sum_{k=0}^{L} k^j x_{t-k} \quad ; \quad j=0,1,\ \ldots,\ P.$$

2. Theorem (Weierstrass): If f is continuous in [a,b] then there exists a polynomial, g, such that $|f(x)-g(x)|<\varepsilon$, for all $x\in[a,b]$, and arbitrarily small $\varepsilon>0$.

In the "Direct Method"[3], for particular P, once L is fixed[4], the $\{\psi_{jt}\}$ series are constructed from $\{x_t\}$, and the α_j's in VII.II.2 are estimated by O.L.S., yielding $\hat{\alpha}_j$, $j=0,1, \ldots, P$. Then, from VII.II.1:

$$\hat{w}_k = \sum_{j=0}^{P} \hat{\alpha}_j k^j \quad ; \quad k=0,1, \ldots, L. \qquad \text{VII.II.3}$$

The (L+2) - dimensional parameter space for model VII.I.2 has been transformed to a space of dimension (P+2) as a result of the restrictions implicit in the use of the polynomial in VII.II.1. Thus, if P is chosen to be substantially less than L, the collinearity problem should be less severe when the $\hat{w}_k$'s are obtained from equation VII.II.3 than if VII.I.2 were estimated directly by O.L.S.

(2) Linear Restrictions

Up to P additional linear (in the α_j's) restrictions may be imposed on any subset of the w_k's, thus constraining the shape of the lag distribution, but in practice only a few types of such restrictions are encountered.

For example, it is common to set one or both of w_0 and w_L to zero, and/or set one or both of $(dw_k/dk)|_0$ and $(dw_k/dk)|_L$ to zero, depending on the desired shape of

---

3. See Fair and Jaffee (1971). Cooper, op.cit., demonstrates the equivalence of the Direct Method and Almon's original method based on Lagrangian interpolation.
4. This limitation always applies to the classical P.A.E., but as is shown in Section VI, it can be relaxed if a Bayesian approach to the problem is adopted.

the estimated lag distribution.[5]

The objective of imposing such constraints is to introduce prior knowledge concerning the shape of f(k) over the range of interest, and hence concerning the pattern of the weights in VII.I.2.. To achieve this objective it is essential that the choice of any such restrictions be made <u>concurrently</u> with the choice of P, since the final outcome depends on both of these factors.

The possibility of imposing constraints of this nature on f(k) certainly adds a degree of flexibility to the P.A.E., but as will be argued in the next part of this Section, it also adds an element of danger.

(3) A Critique

Use of the P.A.E. requires the user to make several selections among what can, in practice, be a large number of possibilities.

In specifying VII.I.2, the values of L and P, and the form and number of any additional linear restrictions must be chosen in such a way as to reflect whatever <u>a priori</u> knowledge of the underlying economic relationship is available. As noted already, any restrictions <u>should</u> be chosen concurrently with the value of P, as both determine the shape of the lag distribution.

---

5. $w_k=\alpha_0+\alpha_1 k+ \ldots + \alpha_P k^P$, and $(dw_k/dk)=\alpha_1+2k\alpha_2+ \ldots +Pk^{P-1}\alpha_P$. Thus, if $w_i=0$, then $\alpha_0+\alpha_1 i+ \ldots +\alpha_P i^P=0$, and this restriction is linear in the α_j's. Further, if $(dw_k/dk)|_j=0$, then $\alpha_1+2j\alpha_2+ \ldots +Pj^{P-1}\alpha_P=0$, which is also linear in the α_j's.

It may be hard to avoid the temptation to test several possible model specifications by considering various combinations of L, P and constraints. This gives rise to the well known problem of pre-testing bias, and if one model specification is to be selected on the basis of the sample evidence, a classicist will probably have to resort to Theil's Theorem[6], and seek to maximize $\overline{R}^2$.

However, some consolation may be found in the fact that by allowing the imposition of linear restrictions, most of the shapes that can be given economic justification can be achieved with polynomials of degree four or less[7], and practical applications are usually limited in this respect. Even so, this may still leave a large number of specifications to be examined.

Mis-specification errors may arise in several ways with the P.A.E.. First, an error arises if in fact the w_k's do not lie on a polynomial of degree P in the interval $[0,L]$. Secondly, if L is understated or if it is overstated by more than P minus the number of linear restrictions, then the model is mis-specified. Thirdly, the (exact) constraints on f(k) should not be imposed unless the *a priori* information is strongly in their favour. This view is supported by Schmidt and Waud (1973), and endorsed by the Monte Carlo evidence in favour of simple lag shapes for the P.A.E. given by McNown (1971). Trivedi (1970) analyses the substantial biases which can arise with inappropriate constraints on f(k), and further evidence is provided by Cohen *et al.* (1973).

---

6. See Theil (1961), pp.212-215.
7. See Fair and Jaffee, op.cit., and Appendix III of this thesis.

Naturally, as model VII.I.2 is generalized to VII.I.1, the number of alternative specifications that might be tested multiplies rapidly.

In short, unless strong a priori knowledge is available and it is adhered to strictly, the "flexibility" of the P.A.E. poses grave dangers for the classical econometrician, for he has no formal means of model-selection, and only too frequently succumbs to gross "data-mining" and the associated pre-testing biases. Accordingly, the validity of inferential statements is distorted, and the final selection of a "preferred" specification must be founded on ad hoc principles.

Certainly, careful methodological practice can reduce the magnitude of these dangers considerably, even for a classical econometrician. A preferred approach is to decide on a limited range of alternative values of L (three or four at most), fix the constraints on f(k) and the value of P to achieve the appropriate shape for f(k), and then estimate the different models corresponding to the different values of L. One model could then be selected on the basis of $\bar{R}^2$ -maximization, as noted earlier.[8]

However, there are still difficulties with this approach. First, as a classicist it is very difficult in practice to limit oneself to even three or four alternative models, especially if the results are not very satisfactory! There is an inherent temptation to "tinker" with the model

8. If the disturbances in VII.I.1 or VII.I.2 are serially correlated then a Cochrane-Orcutt transformation may be used, as is described by Giles (1974). Theil's theorem still holds asymptotically, as is shown by Schmidt (1974).

until it looks "sensible". Just how sensible it looks is often proportional to the number of alternative specifications that have been investigated. In this context, $\bar{R}^2$ is very much a coefficient of determination! In any case, if a particular specification looks sensible a posteriori, why not impose this specification a priori, and estimate its parameters?

Secondly, the classical model-selection procedures are extremely limited. It is meaningless, as a classicist, to attribute probability statements to the alternative model specifications. However, if a Bayesian stance is adopted, then the attribution of such probabilities is not only meaningful and legitimate, but totally in keeping with one's basic philosophy, as is discussed in Chapter I.

The Bayesian approach to econometrics is ideally suited to the incorporation of prior knowledge into an estimation procedure, and to handling model-selection problems. It is thus an ideal framework within which to analyse F.D.L.M.'s.

III. BAYESIAN CONSIDERATIONS

Since w_k is linear in the α_j's, imposing linear restrictions on the w_k's amounts to imposing linear restrictions on the α_j's. For each independent linear restriction placed on the w_k's, one α_j may be eliminated from the expression explaining w_k, and hence from VII.I.2. If P such independent linear restrictions are imposed, the parameter space for VII.I.2 contains only σ and one α_j, and the latter can be chosen to be α_P without loss of generality.

If no restrictions are placed on the lag distribution, the direct Bayesian estimation of VII.I.2 requires the specification of a prior p.d.f. $p(\alpha,\sigma)$, where $\alpha'=(\alpha_0,\alpha_1,\ldots,\alpha_P)$. If P independent linear restrictions are imposed, then this requirement reduces to that of specifying $p(\alpha_P,\sigma)$.

In either case, the specification of the prior p.d.f. is likely to be a difficult task. It is unlikely that one would have enough prior knowledge[9] of the elements of α to be able to do this, since these "Almon coefficients" do not have a direct economic interpretation, so analysis with informative prior p.d.f.'s is likely to be impossible. It should also be noted that if σ and α are assumed independent, and if diffuse[10] prior p.d.f.'s are used, then the Bayes estimates of the w_k's (under quadratic loss) are identical to those in the classical P.A.E..

In their Bayesian version of the P.A.E., ZW consider a special case of VII.I.1, with q=0, and r=2; and they introduce the "sum of weights" parameters, $\theta_j = \sum_{k=0}^{L_j} w_{jk}$; j=1, 2. In their example, $P_j=2$, and two independent linear end-point restrictions are imposed on each lag distribution, so in effect only α_{P_1} and α_{P_2} have to be estimated.

9. If one had sufficient prior information to be able to formulate a prior p.d.f. for the w_k's directly, then one might proceed in the manner suggested by Leamer (1972). However, this possibility seems rather unlikely in practice.

10. That is: $p(\alpha,\sigma) = p(\alpha).p(\sigma)$

$$p(\alpha)d\alpha \propto d\alpha \quad ; \quad -\infty<\alpha_i<\infty$$
$$p(\sigma)d\sigma \propto (d\sigma/\sigma) \quad ; \quad 0<\sigma<\infty$$

ZW do not stress the advantage of introducing the θ_j's - namely, although little prior information about the individual w_k's (or α_j's) may be available, it may be possible to specify a prior p.d.f. for the "sums of weights". Neither do ZW stress the limitations of this approach - if a prior p.d.f. cannot be constructed for the w_k's or α_j's, their method is feasible only if all but one α_j is eliminated from each lag distribution, as is the case, for example, when sufficient independent linear restrictions are imposed on the w_k's.

The prior p.d.f.'s for the θ_j's could be transformed to prior p.d.f.'s for the α_p's, and then no re-parameterization of the model is necessary prior to estimation. However, if the prior information originates with the θ_j's, their posterior p.d.f.'s will be required and these are obtainable most directly if the model is appropriately re-parameterized. Whether or not this is done, the posterior p.d.f.'s for the original w_k's are obtainable by successive transformations, as is discussed in more detail below.

Further, by imposing linear restrictions the dimension of the parameter space is reduced even more than when the P.A.E. is applied without restrictions.[11] The ZW procedure relies critically on the use of linear restrictions, but this is not made explicit by the authors. It is shown in Part (2) of Section IV that if a lag distribution is approximated by a polynomial of degree P_j, then the procedure proposed

11. The parameter space for VII.I.1 is (L+q+2) initially; it is (P+q+2) when the polynomial approximation is introduced; and it is (q+2) if P independent linear restrictions are placed on the w_k's.

by ZW (and generalized here) is not workable unless exactly P_j independent linear homogeneous restrictions are imposed on the lag weights.

This seriously limits the usefulness of the ZW (and our) analysis since, as noted, mis-specification costs are high if restrictions are imposed invalidly, and only a limited range of problems may be analysed by imposing P_j such restrictions. However, as noted above, the economic theory underlying F.D.L.M.'s generally suggests only a few basic distributional shapes, and practical applications with the P.A.E. are limited to these in the literature. Further, these shapes generally can be ensured only by the appropriate use of linear homogeneous restrictions.[12] ZW deal only with exact restrictions in their special case. However, to some extent this inflexibility can be relaxed within the Bayesian framework by allowing different sets of such restrictions to constitute different models, *ceteris paribus*, and assigning prior masses to these models. This possibility is explored further in Section VII, and is applied to some actual data in Chapter VIII.

The ZW estimator reduces to the (restricted) P.A.E. when prior information is diffuse, but potentially the former may be more flexible if some of the advantages of a Bayesian analysis are expoloited. This fact is explored here and in Chapter VIII, and the generalizations which we propose rely heavily on the broad suggestions made by ZW. Essentially, sufficient restrictions are imposed so that the problem is simplified to one of analysing regression models.

---

12. The restrictions noted in Appendix III are in this category.

IV. RE-PARAMETERIZING THE MODEL

We consider model VII.I.2 again, and show that the results extend naturally to VII.I.1. Let L and P be fixed and given[13] and let the only restrictions considered on the w_k's be independent, linear and homogeneous. Then VII.I.2 may be re-parameterized entirely (except for σ) in terms of $\lambda=\lambda(\alpha_P,L)$, where λ is an arbitrary linear homogeneous function of α_P, iff exactly P independent linear homogeneous restrictions are placed on the w_k's.

Since $w_k = (\alpha_0+\alpha_1 k+ \dots +\alpha_P k^P)$ is linear in the α_j's, imposing linear restrictions on the w_k's amounts to imposing linear restrictions on the α_j's. Considering only restrictions which are independent linear homogeneous in the α_j's, for each such restriction on the w_k's, one of the α_j's can be eliminated.

Up to (P+1) <u>independent</u> linear restrictions can be imposed. If any Q(>P+1) linear restrictions are imposed on the w_k's, then (Q-P-1) of them are redundant, and when they are eliminated the shape of lag distribution is specified exactly, though σ^2 is still unknown. (The specification of the shape will depend on which restrictions are eliminated.)

If R(<P+1) such restrictions are imposed, then:

$$w_k = \sum_{i=1}^{P-R+1} \alpha_{P-i+1} f_i(k,L) \quad ; \quad k=0,1, \dots, L.$$

where the form of each f_i depends on P and the restrictions.

---

13. This assumption is relaxed in Sections VI and VII.

If the model is to be re-parameterized in terms of the scalar λ, then:

$$z\lambda = Xw$$

where:

$$w' = (w_0, w_1, \ldots, w_L)$$

$$X = \begin{pmatrix} x_{L+1} & \cdot & \cdot & \cdot & \cdot & \cdot & x_1 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ x_{L+n} & \cdot & \cdot & \cdot & \cdot & \cdot & x_n \end{pmatrix}$$

so,

$$z' = (Xw/\lambda)' = (z_{L+1}, \ldots, z_{L+n})$$

or,

$$z_t = \{\lambda(\alpha_P, L)\}^{-1} \sum_{k=0}^{L} \sum_{i=1}^{P-R+1} \alpha_{P-i+1} f_i(k,L) x_{t-k} \qquad \text{VII.IV.1}$$

Then, even if λ is linear homogeneous in α_P, so that

$$\lambda = \alpha_P . h(L), \qquad \text{VII.IV.2}$$

z_t still depends on the unknown α_j's, so the $\{z_t\}$ series cannot be constructed.

If exactly P such restrictions are imposed,

$$w_k = \alpha_P . f(k,L) \quad ; \quad k=0,1, \ldots, L \qquad \text{VII.IV.3}$$

Then, from VII.IV.1:

$$z_t = \{\lambda(\alpha_P, L)\}^{-1} \sum_{k=0}^{L} \alpha_P . f(k,L) x_{t-k} \qquad \text{VII.IV.4}$$

and α_P may be eliminated from VII.IV.4 iff λ satisfies VII.IV.2, so that:

$$z_t = \{h(L)\}^{-1} \sum_{k=0}^{L} f(k,L) x_{t-k}$$

Clearly the use of θ by ZW emerges as a special case. Once P independent linear homogeneous restrictions are placed on the w_k's, θ is linear homogeneous in α_P. However, other functions also satisfy the above requirements, and a priori information may be available about such quantities.

For example, θ is a special case of a linear combination of the w_k's:

$$\eta = \sum_{k=a}^{b} d_k w_k.$$

Then, from VII.IV.3:

$$\eta = \alpha_P \sum_{k=a}^{b} d_k f(k,L) = \alpha_P . g(L)$$

Another special case of η is if a=b, d_a=1, and all other d_k's are zero. Then the model is re-parameterized in terms of just one of the w_k's. The maximum "height" of the lag distribution is another suitable candidate.

Thus, if a polynomial of degree P is being used, the type of analysis proposed by ZW relies critically on the imposition of P independent linear homogeneous restrictions on the w_k's, and use may be made of prior information about <u>any</u> parameter which is a linear homogeneous function of one of the α_j's, and this can be α_P without loss of generality.

<u>Example</u>: $P=2$; $w_0=w_L=0$.

Under these conditions,

$$w_k = \alpha_0+\alpha_1 k+\alpha_2 k^2 \quad ; \quad k=0,1, \ldots, L.$$

But, $w_0 = 0$ implies that $\alpha_0=0$.

and, $w_L = 0$ implies that $\alpha_0+\alpha_1 L+\alpha_2 L^2=0$.

Thus, $\alpha_1 = -L\alpha_2$;

and, $w_k = -\alpha_2 k(L-k)$.

Also, $f(k,L) = -k(L-k)$;

$h(L) = L(1+L)(1-L)/6$;

$$\theta = \sum_{k=0}^{L} w_k = \alpha_2 L(1+L)(1-L)/6;$$

$$z_t = [6/L(1+L)(1-L)] \sum_{k=0}^{L} (L-k)x_{t-k} \quad ; \quad t=L+1, \ldots, L+n.$$

The above results extend to VII.I.1 if each of the r D.L. variables is treated in turn in the above manner. At this stage, consider all of the L_j's to be fixed and given (although this restriction is relaxed later) and operate on each of the r D.L. variables in the manner now demonstrated for the jth. such variable:

$$w_{jk} = \sum_{m=0}^{P_j} \alpha_{jm} k^m \quad ; \quad k=0,1, \ldots, L_j$$

$$P_j \leq L_j$$

$$j=1,2, \ldots, r.$$

Clearly, the value of P_j may differ from one D.L. variable to another, but in each case P_j independent linear homogeneous restrictions are placed on the w_{jk}'s, and the form of these restrictions may differ from D.L. to another. Thus:

$$w_{jk} = \alpha_{P_j} . f_j(k, L_j) \quad ; \quad k=0,1, \ldots, L_j \quad j=1,2, \ldots, r. \qquad \text{VII.IV.5}$$

where the form of each f_j depends on P_j and the type of restrictions.

Re-parameterization is in terms of:

$$\lambda_j = \alpha_{P_j} . h_j(L_j) \quad ; \quad j=1,2, \ldots, r. \qquad \text{VII.IV.6}$$

where each h_j also depends on P_j and the restrictions. Then:

$$z_{jt} = (\lambda_j^{-1}) \sum_{k=0}^{L_j} w_{jk} x_{jt-k} \quad ; \quad j=1,2, \ldots, r.$$

or,

$$z_{jt} = \{h_j(L_j)\}^{-1} \sum_{k=0}^{L_j} f_j(k, L_j) x_{jt-k} \; ; \; j=1,2, \ldots, r. \qquad \text{VII.IV.7}$$

Again, each $\{z_{jt}\}$ series can be constructed since each z_{jt} is independent of the α_{jm}'s.

V. BAYESIAN ESTIMATION

(1) Serially Independent Errors

Consider the general F.D.L.M., VII.I.1:

$$y_t = \sum_{j=1}^{r} \left(\sum_{k=0}^{L_j} w_{jk} x_{jt-k} \right) + \sum_{i=1}^{q} \phi_i c_{it} + \varepsilon_t, \quad t=L+1, \ldots, L+n.$$

where the ε_t are $NI(0,\sigma^2)$ for all t, and where all of the L_j's are fixed and given.

Under the conditions of the discussion in Section IV, this model may be written:

$$y_t = \sum_{j=1}^{r} \lambda_j z_{jt} + \sum_{i=1}^{q} \phi_i c_{it} + \varepsilon_t \quad ; \quad t=L+1, \ldots, L+n.$$

where λ_j and z_{jt} are defined in VII.IV.6 and VII.IV.7. Given the L_j's, P_j's, constraints and $\{x_{jt}\}$ series, the $\{z_{jt}\}$ series may be constructed. Then:

$$y = (Z,C)\binom{\lambda}{\phi} + \varepsilon$$

or,

$$y = V\xi + \varepsilon \qquad \text{VII.V.1}$$

where:

$$y' = (y_{L+1}, y_{L+2}, \ldots, y_{L+n})$$

$$\varepsilon' = (\varepsilon_{L+1}, \varepsilon_{L+2}, \ldots, \varepsilon_{L+n})$$

$$\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_r)$$

$$\phi' = (\phi_1, \phi_2, \ldots, \phi_q)$$

$$\xi' = (\xi', \phi')$$

$$Z = \begin{pmatrix} z_{1\ L+1} & \cdot & \cdot & \cdot & z_{r\ L+1} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ z_{1\ L+n} & \cdot & \cdot & \cdot & z_{r\ L+n} \end{pmatrix}$$

$$C = \begin{pmatrix} c_{1\ L+1} & \cdot & \cdot & \cdot & c_{q\ L+1} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ c_{1\ L+n} & \cdot & \cdot & \cdot & c_{q\ L+n} \end{pmatrix}$$

$$V = (Z,C) \quad ; \quad \text{rank } (V) = r+q.$$

Further, let:

$$L' = (L_1, L_2, \ldots, L_r)$$

$$P' = (P_1, P_2, \ldots, P_r)$$

$$\mu = \{\mu_j\} \quad ; \quad j=1,2, \ldots, r.$$

where μ_j is the jth. set of P_j independent linear homogeneous restrictions placed on the jth. D.L. variable in VII.I.1.

Given the assumptions on ε, the likelihood function is[14]:

$$p(y|\xi,L,P,\mu,\sigma,V) \propto \sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-V\xi)'(y-V\xi)\} \qquad \text{VII.V.2}$$

14. See Zellner (1971), p.66.

Introduce a suitable[15] prior p.d.f. for the parameters in VII.V.1:

$$p(\xi,\sigma|L,P,\mu) = p(\lambda,\phi,\sigma|L,P,\mu) \qquad \text{VII.V.3}$$

where:

$$-\infty < \lambda_j < \infty \quad ; \quad j=1,2,\ \ldots,\ r$$
$$-\infty < \phi_i < \infty \quad ; \quad i=1,2,\ \ldots,\ q$$
$$0 < \sigma < \infty$$

Applying Bayes' Theorem to VII.V.2 and VII.V.3:

$$p(\xi,\sigma|L,P,\mu,V,y) \propto p(\xi,\sigma|L,P,\mu)p(y|\xi,L,P,\mu,\sigma,V)$$

$$\propto p(\lambda,\phi,\sigma|L,P,\mu).\sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-V\xi)'(y-V\xi)\}$$

Joint inferences about λ and ϕ are based on:

$$p(\xi|L,P,\mu,V,y) = \int_0^\infty p(\xi,\sigma|L,P,\mu,V,y)d\sigma$$

Further, joint inferences about all of the elements of λ are based on[16]:

$$p(\lambda|L,P,\mu,V,y) = \int_{-\infty}^{\infty} p(\xi|L,P,\mu,V,y)d\phi. \qquad \text{VII.V.4}$$

---

15. The exact form of the prior p.d.f. is not important here. However, in some cases one may wish to restrict the ranges of λ and/or ϕ, and this is so in Chapter VIII. Further, a general prior p.d.f. may lead to the need for numerical approximations in subsequent integrations, as is also the case in Chapter VIII.
16. Integration in VII.V.4 is over a q-tuple of elements. Similar notation is used for simplicity elsewhere in this Chapter.

and inferences about all of the elements of ϕ are based on[17]:

$$p(\phi|L,P,\mu,V,y) = \int_{-\infty}^{\infty} p(\xi|L,P,\mu,V,y)d\lambda. \qquad \text{VII.V.5}$$

Further integration of VII.V.4 and VII.V.5 yields the marginal posterior p.d.f.'s for (subsets of) the individual elements of λ and ϕ respectively. These may be used in conjunction with a loss function to obtain M.E.L. estimates of the parameters. For example, as noted in Chapter II, under a quadratic loss function these estimates are the means of the appropriate posterior p.d.f.'s. The lag distributions can be "unscrambled" by using VII.IV.5 and VII.IV.6 to transform the marginal posterior p.d.f. for each λ_j into a marginal posterior p.d.f. for the associated w_{jk}'s:

$$\begin{aligned} p(w_{jk}|L_j,P_j,\mu_j,V,y) &= p(\alpha_{P_j}|L_j,P_j,\mu_j,V,y) \\ &\quad \cdot|f_j(k,L_j)|^{-1} \\ &= p(\lambda_j|L_j,P_j,\mu_j,V,y) \\ &\quad \cdot|f_j(k,L_j)|^{-1}\cdot|h_j(L_j)| \\ &k=0,1,\ \ldots,\ L_j; \quad j=1,2,\ \ldots,\ r \end{aligned}$$

Then M.E.L. estimates, w_{jk}^*, are derived from these marginal posterior p.d.f.'s, taking account of the loss function. This "unscrambling" procedure is <u>not</u> discussed by ZW, though an analogous series of transformations is usually used in

17. Integration in VII.V.5 is over an r-tuple of elements.

the classical Almon estimator.[18]

(2) Serially Correlated Errors

Consider the situation if the disturbances in VII.I.1 are serially correlated according to a first-order Markov process:[19]

$$y_t = \sum_{j=1}^{r} \left(\sum_{k=0}^{L_j} w_{jk} x_{jt-k} \right) + \sum_{i=1}^{q} \phi_i c_{it} + u_t \qquad \text{VII.V.6}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad ;$$

where ε_t is $NI(0,\sigma^2)$ for all t.

Thus:

$$y_t = \rho y_{t-1} + \sum_{j=1}^{r} \left(\sum_{k=0}^{L_j} w_{jk} (x_{jt-k} - \rho x_{jt-k-1}) \right)$$

$$+ \sum_{i=1}^{q} \phi_i (c_{it} - \rho c_{it-1}) + \varepsilon_t .$$

If the appropriate number of independent linear homogeneous restrictions are imposed on each of the r D.L. schemes in VII.V.6, then from Section IV:

$$y = V\xi + u \qquad \text{VII.V.7}$$

$$u = \rho u_{-1} + \varepsilon \qquad \text{VII.V.8}$$

18. See equation VII.II.3 above.
19. Higher-ordered autocorrelation may be handled in an analogous manner. The transformation used to obtain VII.V.9 will then involve longer lags.

where $u' = (u_{L+1}, u_{L+2}, \ldots, u_{L+n})$, and all other symbols are as defined for VII.V.1. Combining VII.V.7 and VII.V.8:

$$y = \rho y_{-1} + (V - \rho V_{-1})\xi + \varepsilon. \qquad \text{VII.V.9}$$

Bayesian analysis of multiple regression models with serially correlated errors is provided by Zellner and Tiao (1964), hereafter ZT, and this analysis may be applied[20] to VII.V.9.

The introduction of the lagged dependent variable in VII.V.9 means that attention must be paid to y_0 (the value of y in period "L"). As noted in Section II of Chapter VI, ZT show that each of the several assumptions that may be made about this "initial condition", leads to the same posterior p.d.f.. Thus, for simplicity, y_0 is assumed fixed and given.

Recalling that $\xi' = (\lambda', \phi')$, the likelihood function for VII.V.9 is:

$$p(y|\lambda,\phi,\sigma,\rho,L,P,\mu,V,y_0)$$
$$\propto \sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}[y-\rho y_{-1}-(V-\rho V_{-1})\xi]'[y-\rho y_{-1}-(V-\rho V_{-1})\xi]\} \qquad \text{VII.V.10}$$

Given the (conditional) prior p.d.f., $p(\lambda,\phi,\sigma,\rho|L,P,\mu)$, Bayes' Theorem yields:

20. This possibility is alluded to, but not pursued by ZW. For an approximate classical counterpart to this analysis see Giles, op.cit..

$$p(\lambda,\phi,\sigma,\rho|L,P,\mu,V,y,y_0) \propto p(\lambda,\phi,\sigma,\rho|L,P,\mu)$$

$$.\sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}[y-\rho y_{-1}-(V-\rho V_{-1})\xi]'[y-\rho y_{-1}-(V-\rho V_{-1})]\} \qquad \text{VII.V.11}$$

ZT analyse only the case where diffuse prior p.d.f.'s are used, but in general this limitation need not be imposed[21] on $p(\lambda,\phi,\sigma,\rho|L,P,\mu)$.

Marginal posterior p.d.f.'s for the individual parameters are obtained by successively integrating VII.V.11. If the Markov process for the errors is to be allowed to be either stable or explosive[22], the range of ρ can be set to $(-\infty,\infty)$ for generality. Then, from VII.V.11:

$$p(\lambda,\phi,\rho|L,P,\mu,V,y,y_0) = \int_0^\infty p(\lambda,\phi,\sigma,\rho|L,P,\mu,V,y,y_0)d\sigma,$$

and,

$$p(\lambda,\phi|L,P,\mu,V,y,y_0) = \int_{-\infty}^\infty p(\lambda,\phi,\rho|L,P,\mu,V,y,y_0)d\rho.$$

Further,

$$p(\lambda|L,P,\mu,V,y,y_0) = \int_{-\infty}^\infty p(\lambda,\phi|L,P,\mu,V,y,y_0)d\phi \qquad \text{VII.V.12}$$

and,

$$p(\phi|L,P,\mu,V,y,y_0) = \int_{-\infty}^\infty p(\lambda,\phi|L,P,\mu,V,y,y_0)d\lambda. \qquad \text{VII.V.13}$$

Then, successive integration of VII.V.12 and VII.V.13 yields the marginal posterior p.d.f.'s for the individual

21. Again, a general prior p.d.f. is likely to lead to non-analytic results, as is the case in Chapter VIII.
22. Restricting ρ to $(-1,+1)$ would ensure a stable Markov process.

λ_j's and ϕ_i's respectively. If desired, the marginal posterior p.d.f. for ρ is obtained as:

$$p(\rho|L,P,\mu,V,y,y_0) = \int_{-\infty}^{\infty} p(\lambda,\phi,\rho|L,P,\mu,V,y,y_0)d\lambda d\phi$$

Finally, the sensitivity of inferences based on specified values of ρ may be ascertained[23] by computing the appropriate conditional posterior p.d.f.'s. Given VII.V.12, the corresponding marginal posterior p.d.f.'s for the w_{jk}'s may be obtained by transformations analogous to those described in Part (1) of this Section, allowing the lag distributions to be "unscrambled".

VI. RANDOM L

So far model VII.I.1 has been analysed under the assumption that all L_j's, P_j's and μ_j's are fixed and known. We now partially relax this assumption and treat the L_j's as discrete random variables.[24] Thus the M.E.L. estimates of the λ_j's and the w_{jk}'s may be made marginal with respect to the L_j's, but are still conditional on the P_j's and μ_j's. Further, it is clear that the analysis in Section V may be viewed as a special case of that presented here. The earlier results may be interpreted as conditional results when the elements of L are random variables, while the results here are marginal with respect to L.

23. See Zellner and Tiao, op.cit., pp.768-770 for some relevant comments. The analysis in Part (1) of this Section is a special case of the analysis here, conditional on $\rho=0$.
24. This possibility is not considered by ZW. Although we consider all of the L_j's to be random, the same type of analysis holds if some of them are not.

As is emphasised in Sections II and III, any P_j must be chosen concurrently with the related μ_j so that the shape of that lag distribution agrees with prior information. So, if the P_j's were treated as discrete random variables, the associated μ_j's would have to be treated in some similar way, but so that only desired combinations of P_j and μ_j arose. It is not clear how this would be done, and we do not pursue this possibility explicitly in this Chapter.

However, some account can be taken of the fact that the P_j's and μ_j's are unlikely to be known with certainty. Taking different combinations of P_j and μ_j defines different model specifications, and here B.P.O. analysis may be used, as is described in Section VII. If two specifications differ only in (some of) the combinations of P_j and μ_j used, then assigning prior masses to the models amounts to assigning prior masses to these combinations of P_j and μ_j. To this extent our analysis may be less prone than is the classical P.A.E. to mis-specification bias resulting from the invalid imposition of additional linear restrictions.

(1) Serially Independent Errors

Consider the situation where the L_j's are discrete random variables (each over some specified range); where we have prior p.m.f.'s for the L_j's; and where each P_j and μ_j is fixed and given. Further, the form of each λ_j is specified.[25]

---

25. That is, it has been decided what function(s) of the w_k's are to be used for re-parameterizing the model, this choice depending on the form of the available a priori information.

The general F.D.L.M. is transformed into the form of equation VII.V.1:

$$y = V\xi + \varepsilon.$$

In the notation of Section V, the elements of L are unknown, but P and μ are fixed and given, and the likelihood function is as in VII.V.2. Consider the following prior p.d.f., based on the assumption that λ, ϕ and σ are independent of L:

$$p(\xi,\sigma|P,\mu) = p(\xi,\sigma|P,\mu,L), \qquad \text{VII.VI.1}$$

and let the prior p.m.f.[26] for L be:

$$p(L|P,\mu) \quad ; \quad L_j = n_j, n_j+1, \ldots, n_j+m_j$$
$$j = 1,2, \ldots, r.$$

In general, it should be plausible to assume that λ, ϕ and σ are distributed independently of L, but such an assumption could be relaxed at the expense of a little computation. Similar assumptions are made by Chetty (1971) and by Palm (1972).

Applying their analysis,[27] Bayes' Theorem yields:

26. When constructing $p(L|P,\mu)$ it may be necessary to allow for dependence between the individual L_j's.
27. In particular, see Palm, op.cit., pp.13-16 where a binomial distributed lag is analysed by Bayesian methods.

$$p(\xi,\sigma|y,L,P,\mu,V)$$

$$\propto p(\xi,\sigma|P,\mu).\sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-V\xi)'(y-V\xi)\},$$

taking account of VII.VI.1.

The posterior p.d.f. is written as <u>conditional</u> on L, since V cannot be constructed unless L is specified. However, this point is resolved below.

Taking account of the randomness of L, <u>marginal</u> inferences about ξ and σ are based on[28]:

$$p(\xi,\sigma|y,P,\mu,V) \propto \sum_L p(L|P,\mu)p(\xi,\sigma|P,\mu,L)$$

$$.\sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-V\xi)'(y-V\xi)\},$$

$$= \sum_L p(\xi,\sigma|y,L,P,\mu,V)p(L|y,\mu,V) \qquad \text{VII.VI.2}$$

while,

$$p(\lambda|y,P,\mu,V) = \int p(\xi,\sigma|y,P,\mu,V)d\phi d\sigma \qquad \text{VII.VI.3}$$

and,

$$p(\phi|y,P,\mu,V) = \int p(\xi,\sigma|y,P,\mu,V)d\lambda d\sigma \qquad \text{VII.VI.4}$$

As before, successive integration of VII.VI.3 and VII.VI.4 yields the marginal posterior p.d.f.'s for (subsets of) the elements of λ and ϕ respectively, and M.E.L. estimates are obtainable once a loss function is specified.

---

28. $\sum_L$ abbreviates the multiple summation: $\sum_{L_1}\sum_{L_2} \ldots \sum_{L_r}$.

The marginal posterior p.m.f. for L is obtained as:

$$p(L|y,P,\mu,V) \propto \int p(L|P,\mu)p(\xi,\sigma|P,\mu,L) \cdot \sigma^{-n}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-V\xi)'(y-V\xi)\}d\xi.d\sigma \qquad \text{VII.VI.5}$$

The various marginal posterior p.m.f.'s for individual L_j's then may be obtained by successive summations of VII.VI.5, and M.E.L. estimates are derivable once a loss function is specified.

Again, the lag distributions can be "unscrambled" in a manner analogous to that in Section V, but account must be taken of the fact that the expressions for w_{jk} and λ_j in VII.IV.5 and VII.IV.6 depend on L_j.

(2) Serially Correlated Errors

The analysis in Part (1) of this Section may be combined with that in Part (2) of Section V. Under the earlier notation, introduce the prior p.d.f.:

$$p(\xi,\sigma,\rho|P,\mu) = p(\xi,\sigma,\rho|P,\mu,L),$$

where ξ, σ and ρ are independent of L; and the prior p.m.f.:

$$p(L|P,\mu).$$

The likelihood function is still VII.V.10, once the model is in the form VII.V.9 with y_0 fixed and given. Then, the posterior p.d.f., conditional on L, is:

$$p(\xi,\sigma,\rho|y,L,P,\mu,V,y_0) \propto p(\xi,\sigma,\rho|P,\mu,L).\ell$$

where ℓ is the likelihood function VII.V.10.

Joint inferences about ξ, ρ and σ (marginal of L) are based on:

$$\begin{aligned} p(\xi,\sigma,\rho|y,P,\xi,V,y_0) &\propto \sum_L p(L|P,\xi)p(\xi,\sigma,\rho|P,\mu,L).\ell \\ &= \sum_L p(\xi,\sigma,\rho|y,L,P,\mu,V,y_0) \\ &\quad .p(L|y,\mu,V,y_0) \end{aligned}$$

Further,

$$p(\lambda|y,P,\mu,V,y_0) = \int p(\xi,\sigma,\rho|y,P,\mu,V,y_0)d\rho.d\phi.d\sigma \qquad \text{VII.VI.6}$$

and,

$$p(\phi|y,P,\mu,V,y_0) = \int p(\xi,\sigma,\rho|y,P,\mu,V,y_0)d\rho.d\lambda.d\sigma \qquad \text{VII.VI.7}$$

Again, successive integration of VII.VI.6 and VII.VI.7 yields the marginal posterior p.d.f.'s for (subsets of) the elements of λ and ϕ respectively, and M.E.L. estimates follow easily once a loss function is nominated.

The marginal posterior p.m.f. for L is obtained as:

$$p(L|y,P,\mu,V,y_0) \propto \int p(L|P,\mu)p(\xi,\sigma,\rho|P,\mu,L).\ell\, d\xi.d\rho.d\sigma$$

Again successive summations of this posterior p.m.f. yields marginal posterior p.m.f.'s for individual L_j's. The "unscrambling" of the lag distributions follows the method used when the errors are serially independent. Thus, whether L is fixed or random, and whether or not the errors are independent or correlated, there are few conceptual difficulties with a Bayesian analysis. However, computational costs are likely to be high if numerical integration must be used to derive the posterior p.d.f.'s and p.m.f.'s, especially if more than three or four parameters are involved.

VII. MODEL COMPARISONS

(1) Computation of the B.P.O.

Alternative specifications of VII.I.1 (or VII.V.6) arise frequently with the P.A.E., and although this problem is not analysed by ZW, they point out that Bayesian methods may be helpful here. Different specified values of some or all of the L_j's, P_j's or μ_j's lead to different models, and the B.P.O. analysis described in Section V of Chapter II provides a formal and practical means of model discrimination. In general, no such tool is available with the classical P.A.E., but recall the comments concerning the use of $\bar{R}^2$ in Section II above.

For simplicity the discussion below is in terms of just two alternative specifications of the model VII.V.6, but it applies to any model space of finite dimension. However many specifications are considered, each should be based on the same n observations if the B.P.O. are to be able to discriminate.[29] This is especially pertinent if the different specifications arise (at least partly) by

29. See Maddala (1971), p.17.

varying some L_j's.

Consider two model specifications:

$$M_v: y_t = \sum_{j=1}^{r_v} \left(\sum_{k=0}^{L_{vj}} (w_v)_{jk} (x_v)_{jt-k} \right) + \sum_{i=1}^{q_v} (\phi_v)_i (c_v)_{it} + u_{vt}$$

$$u_{vt} = \rho_v u_{vt-1} + \varepsilon_{vt} \quad ; \quad v=1,2. \qquad \text{VII.VII.1}$$

where: ε_{vt} is $NI(0,\sigma_v^2)$ for all t; v=1,2.

$$(w_v)_{jk} = \sum_{m=0}^{P_{vj}} (\alpha_v)_{jm} k^m \quad ; \quad k=0,1, \ldots, L_{vj}$$

$$j=1,2, \ldots, r_v$$

$$v=1,2.$$

P_{vj} independent linear homogeneous restrictions, μ_{vj}, are placed on the jth. lag distribution in VII.VII.1, for all j, and each lag is re-parameterized in terms of:

$$(\lambda_v)_j = (\alpha_v)_{P_{vj}} h_{vj}(L_{vj}) \quad ; \quad j=1,2, \ldots, r_v.$$

$$v=1,2.$$

Let:

$$y' = (y_{L+1}, y_{L+2}, \ldots, y_{L+n})$$

$$\varepsilon_v' = (\varepsilon_{v\ L+1}, \varepsilon_{v\ L+2}, \ldots, \varepsilon_{v\ L+n})$$

$$\lambda_v' = ((\lambda_v)_1, (\lambda_v)_2, \ldots, (\lambda_v)_{r_v})$$

$$\phi_v' = ((\phi_v)_1, (\phi_v)_2, \ldots, (\phi_v)_{q_v})$$

$$L_v' = ((L_v)_1, (L_v)_2, \ldots, (L_v)_{r_v})$$

$$P'_v = ((P_v)_1, (P_v)_2, \ldots, (P_v)_{r_v})$$

$$\mu'_v = ((\mu_v)_1, (\mu_v)_2, \ldots, (\mu_v)_{r_v})$$

for $v=1,2$.

Two cases are now analysed. First, consider the situation when L_1 and L_2 are fixed, and when at least part of the difference between M_1 and M_2 results from differences in some of the elements of L_1 and L_2. Then, the following are <u>given</u>:

(i) $p(M_v)$; $p(\lambda_v, \phi_v, \sigma_v, \rho_v | L_v, P_v, \mu_v)$; $v=1,2$.

(ii) L_v, P_v, μ_v ; $v=1,2$.

(iii) The <u>forms</u> of the parameters of λ_v ; $v=1,2$.

(iv) y_0

The B.P.O. are[30]:

$$\{p(M_1|y,y_0)/p(M_2|y,y_0)\} = \{p(M_1)/p(M_2)\}\{Q_1/Q_2\}$$

where:

$$Q_v = \int p(\lambda_v, \phi_v, \sigma_v, \rho_v | L_v, P_v, \mu_v) \cdot p(y|\lambda_v, \phi_v, \sigma_v, \rho_v, L_v, \mu_v, V, y_0) d\lambda_v . d\phi_v . d\sigma_v . d\rho_v \quad ; \quad v=1,2. \qquad \text{VII.VII.2}$$

30. See equations II.II.1 and II.V.2.

In the special case where the errors are serially independent and VII.I.1 holds, then VII.VII.2 reduces to:

$$Q_v' = \int p(\lambda_v, \phi_v, \sigma_v | L_v, P_v, \mu_v) p(y | \lambda_v, \phi_v, \sigma_v, L_v, \mu_v, V) \cdot d\lambda_v \cdot d\phi_v \cdot d\sigma_v \quad ; \quad v=1,2.$$

Secondly, if L_1 and L_2 are vectors of discrete random variables, then the differences between M_1 and M_2 arise either from the choice of variables, or from the choices of P_1, P_2, μ_1 and μ_2. Then, the following are given:

(i) $p(M_v)$; $p(\lambda_v, \phi_v, \sigma_v, \rho_v | P_v, \mu_v)$; $p(L_v | P_v, \mu_v)$; $v=1,2$.

(ii) P_v, μ_v ; $v=1,2$.

(iii) The forms of the parameters of λ_v ; $v=1,2$.

(iv) y_0.

The B.P.O. are:

$$\{p(M_1 | y, y_0)/p(M_2 | y, y_0)\} = \{p(M_1)/p(M_2)\}\{R_1/R_2\}$$

where:

$$R_v = \int \sum_L p(\lambda_v, \phi_v, \sigma_v, \rho_v | P_v, \mu_v) p(L_v | P_v, \mu_v) \cdot p(y | \lambda_v, \phi_v, \sigma_v, \rho_v, L_v, \mu_v, V, y_0) d\lambda_v \cdot d\phi_v \cdot d\sigma_v \cdot d\rho_v ; \quad v=1,2. \qquad \text{VII.VII.3}$$

If the errors are serially independent, then VII.VII.3

reduces to:

$$R'_v = \int\sum_L p(\lambda_v,\phi_v,\sigma_v \; P_v,\mu_v)\, p(L_v|P_v,\mu_v)$$

$$\cdot p(y|\lambda_v,\phi_v,\sigma_v,L_v,\mu_v,V)\, d\lambda_v . d\phi_v . d\sigma_v \quad ; \quad v=1,2.$$

If the B.P.O. are converted to posterior probabilities, then care must be taken that the model space is fully specified.[31] Finally, note that in the first of the two cases just analysed where L_1 and L_2 are fixed, if the two models differ only in the value of L, then the B.P.O. for M_1 and M_2 reduce to the B.P.O. for L_1 and L_2.

(2) A Qualification

Some comment must be made concerning the computation of the B.P.O. in VII.VII.2 and VII.VII.3 when the disturbances follow a first-order Markov process.

The data densities appearing in the expressions for Q_v and R_v are as in VII.V.10, and thus are based on the transformed models of the form VII.V.9. As was noted and exploited above in the context of M.E.L. estimation, once the original models are transformed to the form VII.V.9, the resulting disturbance terms are spherical, and in this case B.P.O. analysis is straightforward. This approach to the computation of B.P.O. for models involving a first-order Markov process is also adopted by Geisel and by Courville and Geisel.[32] The "cost" of such transformations

31. See Zellner, op.cit., p.316, concerning the conversion of B.P.O. to posterior probabilities.
32. See Geisel (1970), pp.82-86; Courville and Geisel (1971), p.14.

lies in the need to introduce some assumption about the "initial conditions" for the models.

A more general discussion of B.P.O. analysis for models with non-spherical disturbances (of which we have here a special case) is given by Lempers (1971), Gaver and Geisel (1973), and Gaver (1974).

The method adopted in this Section (and in Chapter VIII) may be seen to be a very close approximation to that proposed by Gaver and Geisel. They analyse the models with first-order Markov processes in a way that constitutes a Bayesian analogue to the Generalized Least Squares (G.L.S.) analysis in classical estimation theory. By following ZT, we are effectively working with a Bayesian analogue to the well-known Cochrane-Orcutt transformation.[33] Using this transformation in classical analysis is almost equivalent[34] to using the G.L.S. estimator, the only difference arising from the treatment of the first observation[35] on y and V. By the same token, our B.P.O. analysis in Part (1) of this Section is equivalent to that proposed by Gaver and Geisel if account is taken of the different ways in which the two methods treat the "initial conditions" of the models.

VIII. CONCLUSIONS

This Chapter considers a Bayesian interpretation of the classical polynomial approximation estimator for the

---

33. See Cochrane and Orcutt (1949).
34. See Johnston (1972), pp.259-262.
35. See Kadiyala (1968). Our method involves forming $(y_1 - \rho y_0)$ and $(V_1 - \rho V_0)$ while a G.L.S. approach results in $(\sqrt{1-\rho^2})y_1$ and $(\sqrt{1-\rho^2})V_1$, respectively.

general finite distributed lag model.

The approach adopted here incorporates several generalizations of that proposed for a particular case by Zellner and Williams, most of these generalizations being suggested in principle, but not pursued by those authors. Further, the limitations of this approach are made explicit.

Particular attention is given to: alternative ways of incorporating prior information; treating the extremities of the lag distributions as unknown discrete random variables; using Bayesian Posterior Odds for model discrimination; and allowing for serially correlated errors.

Although numerical approximations are likely to be needed when integrating the various density functions, thus adding to computational expense, the result is a Bayesian procedure which displays all of the flexibility of its classical counterpart, yet has many other desirable features. These features are explored in the practical application discussed in the next Chapter.

CHAPTER VIII

BAYESIAN INFERENCE AND THE RESTRICTED ALMON ESTIMATOR. PART II : AN APPLICATION TO CURRENT PAYMENTS FOR NEW ZEALAND IMPORTS.

I. INTRODUCTION

This Chapter discusses a practical application of the Bayesian estimation procedure described in Chapter VII, this procedure generalizing the Bayesian version of the Almon estimator presented for a special case by Zellner and Williams (1973).

The economic relationship studied here is essentially that postulated by Deane et al. (1972) and Deane and Lumsden (1972) to explain current payments for c.i.f. imports into New Zealand.

Those authors have modified the form of this equation substantially in more recent work, but the earlier simpler form is better suited to our present purpose: namely, to illustrate the features of a new estimation procedure.[1]

An important feature of the relationship is that it involves few unknown parameters, even if the disturbances are assumed to be first-order serially correlated, and this is important in view of the need to use numerical integration routines in parts of this study. These routines were written by the author and are similar to those described by Geisel (1970; pp. 118-119) and Zellner (1971; pp. 409-414).

---

1. See Deane et al., op.cit., for a discussion of the economic theory underlying the basic relationship.

II. DATA

The equation to be estimated measures the lag relationship between c.i.f. imports into New Zealand, and the payments for these goods. The lag is determined by credit offered by the overseas supplier, and by domestic finance available to the importer. Goods imported under no-remittance licences are excluded since the payments for such goods do not appear in the Overseas Exchange Transactions (O.E.T.) accounts of the Reserve Bank of New Zealand.

The relationship is:

$$y_t = \sum_{k=0}^{L} w_k I_{t-k} + \beta D_t + u_t \qquad \text{VIII.II.1}$$

where: y = current payments for c.i.f. imports (O.E.T. basis); \$m. <u>Source</u>: O.E.T. Section, Economic Department,Reserve Bank of New Zealand.

I = total c.i.f. imports; less exogenous items (aircraft, ships, railway equipment, arms of war, etc.); less imports authorised under no-remittance licences; \$m. <u>Source</u>: Department of Statistics <u>Monthly Abstract of Statistics</u> (External Trade Supplement); and Research Section, Economic Department, Reserve Bank of New Zealand.

D = dummy variable to represent official monetary policy with respect to trading banks' advances. <u>Source</u>: Research Section, Economic Department, Reserve Bank of New Zealand.

The raw data appear in Appendix IV.

In some parts of the study the u_t are assumed to be $NID(0,\sigma_1^2)$, while in other parts first-order serial correlation is assumed:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \text{VIII.II.2}$$

with ε_t $NID(0,\sigma_2^2)$, for all t.

In general, the data are consistent with those used by Deane et al., but in some respects our sample differs slightly from theirs. First, we use quarterly data for the period 1961(3) to 1972(1) as the basic sample, this period being longer than that used by Deane et al.. Secondly, the series for c.i.f. imports has been revised slightly since the study by Deane et al., and we use the revised series.[2] Thirdly, the dummy variable, D is re-defined to be minus one times the variable constructed by Deane et al.. Thus, β is expected to be positive, a priori, not negative, and this simplifies the construction of informative prior p.d.f.'s for this parameter. However, the absolute magnitudes of β and its estimates are unchanged.

The construction of the $\{D_t\}$ series implies a linear relationship between periods of "tight", "moderate" and "easy" monetary policy, and this construction is accepted for the purposes of this study. However, other constructions are possible and could be tested by menas of B.P.O. analysis. The present construction of $\{D_t\}$ could lead to an errors-in-variables problem, but this possibility

---

2. These revisions are of a very minor nature and have been found to have a negligible effect on the results of this study.

is not pursued here.[3]

III. SEASONAL ADJUSTMENT

Deane et al. allow for seasonality in the data by including three additive seasonal dummy variables and a constant term.[4] This is equivalent to a special case of the M.E.L. procedure suggested in Chapter IV, with D omitted from IV.II.1 when adjusting each series.[5] The latter is also equivalent to adding seasonal dummy variables to VIII.II.1 and assigning them diffuse prior p.d.f.'s for M.E.L. estimation. However, this would add substantially to computational burden if any of the other parameters in VIII.II.1 were assigned informative prior p.d.f.'s.

Accordingly, we seasonally adjust the data prior to estimating VIII.II.1, using the method of Chapter IV. An attractive side-benefit of prior adjustment is that the number of unknown parameters in VIII.II.1 is kept small, which is important if numerical integration methods must be used when estimating the parameters in this relationship.

In applying the analysis of Chapter IV, two important approximations are made. First, we abstract from the problem of simultaneity bias (as we do elsewhere in this study) and so we use the techniques intended for individual time-series, even though these series then appear in a regression. This avoids the additional computational

---

3. Zellner (1971), Chapter 5, provides a Bayesian analysis of this problem.
4. Equivalently, four additive seasonal dummy variables could be used.
5. See Lovell (1963).

problems associated with Bayesian estimation and model-comparisons in a simultaneous system.

Secondly, the use of B.P.O. analysis to determine the form of the systematic components, as discussed in Section VI of Chapter IV, is complicated by our a priori ingorance about the parameters in IV.II.1.

As yet there is no fully satisfactory way of applying B.P.O. analysis to models with different numbers of parameters when the prior p.d.f.'s are diffuse.[6] The only possibility is that suggested by Geisel[7], and although it is based on what he admits are strong and unrealistic assumptions, we use his result that under these assumptions the M.E.L. rule leads to choosing the model which maximizes $\overline{R}^2$. This appears to be the only Bayesian justification for Theil's Theorem, mentioned in Chapter VII.

Thus when testing alternative D matrices of the form IV.IV.1 in equation IV.II.1, we use an "$\overline{R}^2$-delete" procedure, which amounts to retaining a particular power of T in the D matrix only if the estimated t-value on its coefficient exceeds unity.[8] The final series are then scaled to have the same sample means as their unadjusted counterparts so that a priori information is not distorted. Admittedly, the Bayesian justification for this procedure is rather weak.

However, the results of the study are fairly insensitive to the choice of method for seasonal adjustment. Some tests were made with series adjusted by the X-11 variant of

---

6. In particular, see the excellent survey by Gaver and Geisel (1974), pp.66-72.
7. See Geisel (1970), pp.41-44.
8. See Haitovsky (1969).

the Census Method II, and by omitting the matrix D from IV.II.1. We did not undertake a full B.P.O. analysis, but the results obtained were almost identical with those tabulated below.[9]

Thus, although the adjustment method used here is a crude approximation to that discussed in Chapter IV, it seems that in this instance the results are very similar to those arising with other well known and widely used non-Bayesian methods of adjustment.

IV. ESTIMATION

In this Section we apply to VIII.II.1 the theoretical analysis developed in the last Chapter. We consider three different specifications for the lag distribution in this model:

Specification 1: $P = 2; \quad w_L = (dw_k/dk)|_L = 0$

Specification 2: $P = 2; \quad w_L = (dw_k/dk)|_0 = 0$

Specification 3: $P = 3; \quad w_L = (dw_k/dk)|_0 = (dw_k/dk)|_L = 0$

The lag shapes implied by these specifications appear in Appendix III, each specification being such that P independent linear homogeneous restrictions are imposed upon the w_k's, when the latter are approximated by the ordinates of a polynomial of degree P.

If $\theta = \sum_{k=0}^{L} w_k$, then VIII.II.1 may be written:

9. The simple correlations between the M.E.L. and Census seasonally adjusted versions of y and I are 0.9979 and 0.9949 respectively. See Appendix V.

$$y_t = \theta z_t + \beta D_t + u_t, \qquad \text{VIII.VI.1}$$

where the expressions for θ, w_k and z_t are:

Specification 1:

$$\theta = \alpha_2 L(L+1)(2L+1)/6$$

$$w_k = \alpha_2 (k-L)^2$$

$$z_t = [6/L(L+1)(2L+1)] \sum_{k=0}^{L} (k-L)^2 I_{t-k}$$

Specification 2:

$$\theta = -\alpha_2 L(L+1)(4L-1)/6$$

$$w_k = \alpha_2 (k-L)(k+L)$$

$$z_t = [-6/L(L+1)(4L-1)] \sum_{k=0}^{L} (k-L)(k+L) I_{t-k}$$

Specification 3:

$$\theta = \alpha_3 L^3 (L+1)/4$$

$$w_k = \alpha_3 (L^3 - 3Lk^2 + 2k^3)/2$$

$$z_t = [4/L^3(L+1)] \sum_{k=0}^{L} [(L^3 - 3Lk^2 + 2k^3)/2] I_{t-k}$$

Letting $\xi' = (\theta, \beta)$ and $V = (z, D)$, then VIII.IV.1 may be re-written as:

$$y = V\xi + u \qquad \text{VIII.IV.2}$$

where u may be serially independent, or first-order serially correlated. The $\{z_t\}$ series (and hence V) can be constructed only if a value is assigned to L. In this study we consider only the four values L = 3,4,5,6. In conjunction with the above three specifications, this gives rise to twelve different models for the basic relationship VIII.II.1 under a particular assumption about the disturbances.

Models M_1 to M_4 are derived from specification 1 in conjunction with L = 3,4,5 and 6 respectively. Models M_5 to M_8 are derived from specification 2 in conjunction with L = 3,4,5 and 6 respectively. Finally, Models M_9 to M_{12} are derived from specification 3 in conjunction with L = 3,4, 5 and 6 respectively.

The Bayesian estimation and model-comparison procedures discussed in Chapter VII are applied to these twelve models via equation VIII.IV.2, making use of the available prior information concerning the unknown parameters.

V. PRIOR KNOWLEDGE

For the purposes of this study, we accept VIII.II.1, and hence VIII.IV.2, as the basic relationship. The study could be extended in several ways, each making use of B.P.O. analysis. Alternative and/or additional explanatory variables could be tested; different definitions of the data series could be compared[10] (in particular, one could compare different methods of seasonal adjustment); and other forms of serial correlation might be investigated.

Such extensions would greatly compound an already expensive experiment, so while acknowledging such possibilities, we concentrate on those features of the estimation procedure which in our view make it more versatile than its classical counterpart.

The three specifications defined in Section IV are chosen to reflect the prior belief that the lag distribution

10. For example, see Geisel, op.cit., Chapter 4.

in VIII.II.1 should reflect a declining impact of I on y over time. Specification 1 with L=4 (i.e. model 2, M_2) is that chosen by Deane et al..

(1) The Model Space

A priori indifference between the twelve models leads to the prior p.m.f.:

$$p(M_i) = (1/12) \quad ; \quad i=1,2, \ldots, 12.$$

When L is treated as a discrete random variable, the model space collapses to three elements, these being the three specifications of Section IV. In that case:

$$p(S_i) = (1/3) \quad ; \quad i=1,2,3.$$

(2) The Parameter Space

When L is treated as a discrete random variable, prior ignorance leads to the prior p.m.f.:

$$p(L) = (1/4) \quad ; \quad L=3,4,5,6.$$

The "sum of weights" parameter, θ, is expected to be close to unity if L is properly specified, so when proper prior information is being incorporated, θ is restricted to the interval $[0.9, 1.1]$, with $\mathcal{E}(\theta) = 1.0$. That is, our prior feelings discount the possibilitiy of any substantial "leakages".

The parameter β is expected to be non-negative, since the stronger the restrictions imposed on trading banks' advances the <u>less</u> payments are expected to be made in that period. The results obtained by Deane <u>et al</u>., and a consideration of our own modifications of the data suggest that β should lie in $[0.0, 6.0]$. This restriction is imposed whenever proper prior information is used.

We are <u>a priori</u> ignorant concerning the error variance, σ^2, except that σ is finite and positive. Thus a diffuse Jeffreys' prior p.d.f. is always used for σ:

$$p(\sigma) \propto \sigma^{-1} \quad ; \quad 0 < \sigma < \infty.$$

When the distrubances are assumed to be first-order serially correlated, ρ is taken to be positive and the correlation scheme stable, so ρ is restricted to the interval $[0.0, 0.99]$, whenever proper prior information is being used. This range is suggested by the low values of the Durbin-Watson statistic in Table VIII.VIII.1 below.

In all cases, the parameters are taken to be mutually independent, so that

$$p(\theta,\beta,\sigma,\rho,L) = p(\theta).p(\beta).p(\sigma).p(\rho).p(L),$$

though of course $p(\rho)$ and $p(L)$ are not relevant to certain parts of the study. When Jeffreys' diffuse prior is used as a "reference" prior p.d.f., we have:

$$p(\theta) \mathrel{\dot{\propto}} \text{const.} \quad ; \quad -\infty < \theta < \infty$$
$$p(\beta) \mathrel{\dot{\propto}} \text{const.} \quad ; \quad -\infty < \beta < \infty$$
$$p(\rho) \mathrel{\dot{\propto}} \text{const.} \quad ; \quad -\infty < \rho < \infty$$

$$p(\sigma) \doteq \sigma^{-1}$$

where $\doteq$ denotes that prior information is "negligible" rather than "zero".

Two informative joint prior p.d.f.'s are tested, these differing from each other only in the construction of $p(\theta)$. In each case the following marginal prior p.d.f.'s are used:

$$p(\sigma) \propto \sigma^{-1} \quad ; \quad 0 < \sigma < \infty$$

$$p(\beta) \propto [(\beta/6)(1-\beta/6)]^{0.75} \quad ; \quad 0 \leq \beta \leq 6.0$$

and, where appropriate:

$$p(\rho) \propto [(\rho/0.99)(1-\rho/0.99)]^{1.5628} \quad ; \quad 0 \leq \rho \leq 0.99$$

$$p(L) = (1/4) \quad ; \quad L=3,4,5,6.$$

That is, "loose" beta p.d.f.'s are used for β and ρ, having prior means of 3.0 and 0.5 respectively, and prior variances of 2.0 and 0.04 respectively. The prior p.m.f. for L is uniform over its range, with mean 4.5 and variance 1.25.

The two alternative (beta) prior p.d.f.'s for θ differ only in the prior variance, thus giving rise to a "loose" and "tight" joint prior p.d.f. for the parameters:

$$p_1(\theta) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{0.5} \quad ; \quad 0.9\leq\theta\leq1.1$$

$$p_2(\theta) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{48.5} \quad ; \quad 0.9\leq\theta\leq1.1$$

The prior means of p_1 and p_2 are both 1.0, and the prior variances are 0.0025 and 0.0001 respectively.

A comparison of the results obtained with the "loose" and "tight" joint prior p.d.f.'s, and the "diffuse" (or "reference") joint prior p.d.f. permits an evaluation of the extent to which the use of proper prior information affects the parameter estimates, and the degree to which the sample information (in the likelihood) is dominated.

VI. POSTERIOR DISTRIBUTIONS

The likelihood function for the sample depends upon the value of L, the type of restrictions imposed upon the w_k's (and hence depends upon the choice of <u>specification</u>), and upon whether the disturbances are serially independent or correlated.

The study falls into four parts:

(i) L fixed and known; errors serially independent.

(ii) L fixed and known; errors first-order serially correlated.

(iii) L random; errors serially independent.

(iv) L random; errors first-order serially correlated.

As noted in Chapter VII, the first two parts of the study may be interpreted as special cases of the second two parts: namely when L is random but all of the analysis is <u>conditional</u> on particular values of this parameter, rather than being <u>marginal</u> with respect to L. Similarly, the analysis under the assumption of serial independence may be viewed as conditional (ρ=0) analysis under the assumption of first-order serial correlation.

The likelihood function for VIII.IV.2 is the data density, viewed as a function of the parameters. From VII.V.2, this is:

$$p(y|\theta,\beta,\sigma_1,L,S) \propto \sigma_1^{-n}\exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}$$

if the errors are serially independent; or, from equation VII.V.10:

$$p(y|\theta,\beta,\rho,\sigma_2,L,S,y_0) \propto \sigma_2^{-n}\exp\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n[y_t-\rho y_{t-1} -\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}$$

if the errors are first-order serially correlated. Note that S denotes the <u>specification</u> (so it reflects both P and the restrictions), y_0 is assumed to be fixed and given, and n=43. Consider the various posterior p.d.f.'s and p.m.f.'s for the four parts of the study. These p.d.f's and p.m.f.'s are obtained by making repeated use of equations VII.V.4, VII.V.5, VII.V.11, VII.V.12, VII.V.13, VII.VI.2, VII.VI.3, VII.VI.4, VII.VI.5, VII.VI.6 and VII.VI.7.

(1) <u>L Fixed and Known; Errors Serially Independent</u>

(a) <u>"Diffuse" prior</u>

$$p(\theta,\beta,\sigma_1|y,L,S) \dot{\propto} \sigma_1^{-(n+1)}\exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}$$

and,

$$p(\theta,\beta|y,L,S) \dot{\propto} \{\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}},$$

which is a bivariate Student-t distribution.

(b) "Loose" prior

$$p(\theta,\beta,\sigma_1|y,L,S) \propto \{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{0.5}$$

$$.\{(\beta/5)(1-\beta/6)\}^{0.75}.\sigma_1^{-(n+1)}$$

$$.\exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}$$

and,

$$p(\theta,\beta|y,L,S) \propto \{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{0.5}$$

$$\{(\beta/0.6)(1-\beta/6)\}^{0.75}\{\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}$$

(c) "Tight" prior

$$p(\theta,\beta,\sigma_1|y,L,S) \propto \{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{48.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\sigma_1^{-(n+1)}$$

$$.\exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}$$

and,

$$p(\theta,\beta|y,L,S) \propto \{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{48.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}$$

(2) L Fixed and Known; Errors Serially Correlated

(a) "Diffuse" prior

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,L,S) \dot{\propto} \sigma_2^{-(n+1)}\exp\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n[y_t-\rho y_{t-1}$$

$$-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}$$

and,

$$p(\theta,\beta,\rho|y,y_0,L,S) \,\dot{\propto}\, \{\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1}) -\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}$$

and[11],

$$p(\theta,\beta|y,y_0,L,S) = \int_{-\infty}^{\infty} p(\theta,\beta,\rho|y,y_0,L,S)\,d\rho.$$

(b) "Loose" prior

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,L,S) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{0.5} .\exp\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1}) -\beta(D_t-\rho D_{t-1})]^2\}$$

and,

$$p(\theta,\beta,\rho|y,y_0,L,S) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{0.5} .\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99) .(1-\rho/0.99)\}^{1.5628}\{\sum_1^n [y_t-\rho y_{t-1} -\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}$$

and,

$$p(\theta,\beta|y,y_0,L,S) = \int_0^{0.99} p(\theta,\beta,\rho|y,y_0,L,S)\,d\rho.$$

---

11. The ranges of any indefinite integrals are as discussed in Section V above.

(c) "Tight" prior

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,L,S) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{48.5}$$
$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99)$$
$$.(1-\rho/0.99)\}^{1.5628}\sigma_2^{-(n+1)}.$$
$$.\exp\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n[y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1})$$
$$-\beta(D_t-\rho D_{t-1})]^2\}$$

and,

$$p(\theta,\beta,\rho|y,y_0,L,S) \propto \{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{48.5}$$
$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99)$$
$$.(1-\rho/0.99)\}^{1.5628}\{\sum_1^n[y_t-\rho y_{t-1}$$
$$-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}$$

and,

$$p(\theta,\beta|y,y_0,L,S) = \int_0^{0.99} p(\theta,\beta,\rho|y,y_0,L,S)d\rho.$$

(3) L Random; Errors Serially Independent

(a) "Diffuse" prior

$$p(\theta,\beta,\sigma_1|y,S) \dot{\propto} \sum_{L=3}^{6}\tfrac{1}{4}\sigma_1^{-(n+1)}\exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}$$

$$p(\theta,\beta|y,S) \dot{\propto} \sum_{L=3}^{6}\tfrac{1}{4}\{\sum_1^n(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}$$

and,

$$p(L|y,S) \;\dot{\propto}\; \int_{-\infty}^{\infty} \tfrac{1}{4}\{\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}\, d\theta . d\beta$$

(b) "Loose" prior

$$p(\theta,\beta,\sigma_1|y,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\sigma_1^{-(n+1)} \exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}$$

$$.\{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{0.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$p(\theta,\beta|y,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\{\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}} . \{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$.\{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{0.5}$$

and,

$$p(L|y,S) \propto \int \tfrac{1}{4}\{\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}\{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$.\{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{0.5} d\theta . d\beta$$

(c) "Tight" prior

$$p(\theta,\beta,\sigma_1|y,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\sigma_1^{-(n+1)} \exp\{-\tfrac{1}{2}\sigma_1^{-2}\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}$$

$$.\{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{48.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$p(\theta,\beta|y,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\{\sum_{1}^{n}(y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}} . \{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$.\{[(\theta-0.9)/0.2]\,[1-(\theta-0.9)/0.2]\}^{48.5}$$

$$p(L|y,S) \propto \int \tfrac{1}{4}\{\sum_1^n (y_t-\theta z_t-\beta D_t)^2\}^{-\frac{n}{2}}.\{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$.\{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{48.5}d\theta.d\beta$$

(4) L Random; Errors Serially Correlated

(a) "Diffuse" prior

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,S) \,\dot{\propto}\, \sum_{L=3}^{6} \tfrac{1}{4}\sigma_2^{-(n+1)}\exp.\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n [y_t-\rho y_{t-1}$$

$$-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}$$

$$p(\theta,\beta,\rho|y,y_0,S) \,\dot{\propto}\, \sum_{L=3}^{6} \tfrac{1}{4}\{\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1})$$

$$-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}$$

$$p(\theta,\beta|y,y_0,S) = \int_{-\infty}^{\infty} p(\theta,\beta,\rho|y,y_0,S)\,d\rho.$$

and,

$$p(L|y,y_0,S) \,\dot{\propto}\, \int_{-\infty}^{\infty} \tfrac{1}{4}\{\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1})$$

$$-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}d\theta.d\beta.d\rho$$

(b) "Loose" prior

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\sigma_2^{-(n+1)}\exp.\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n [y_t-\rho y_{t-1}$$

$$-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}$$

$$.\{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{0.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99)$$

$$.(1-\rho/0.99)\}^{1.5628}$$

$$p(\theta,\beta|y,y_0,S) = \int_0^{0.99} p(\theta,\beta,\rho|y,y_0,S)d\rho$$

and,

$$p(L|y,y_0,S) \propto \int \tfrac{1}{4}\{\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}$$

$$.\{[(\theta-0.99)/0.2][1-(\theta-0.99)/0.2]\}^{0.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99)$$

$$.(1-\rho/0.99)^{1.5628}d\theta.d\beta.d\rho$$

(c) <u>"Tight" prior</u>

$$p(\theta,\beta,\rho,\sigma_2|y,y_0,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\sigma_2^{-(n+1)}\exp.\{-\tfrac{1}{2}\sigma_2^{-2}\sum_1^n [y_t-\rho y_{t-1}$$

$$-\theta(z_t-\rho z_{t-1})-\beta(D_t-\rho D_{t-1})]^2\}$$

$$.\{[(\theta-0.9)/0.2][1-(\theta-0.9)/0.2]\}^{48.5}$$

$$.\{(\beta/6)(1-\beta/6)\}^{0.75}\{(\rho/0.99)$$

$$.(1-\rho/0.99)\}^{1.5628}$$

$$p(\theta,\beta,\rho|y,y_0,S) \propto \sum_{L=3}^{6} \tfrac{1}{4}\{\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1})$$

$$-\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}\{[(\theta-0.9)/0.2]$$

$$.[1-(\theta-0.9)/0.2]\}^{48.5}\{(\beta/6)(1-\beta/6)\}^{0.75}$$

$$.\{(\rho/0.99)(1-\rho/0.99)\}^{1.5628}$$

$$p(\theta,\beta|y,y_0,S) = \int_0^{0.99} p(\theta,\beta,\rho|y,y_0,S)d\rho.$$

and,

$$p(L|y,y_0,S) \propto \int\int\int \{\textstyle\sum_1^n [y_t-\rho y_{t-1}-\theta(z_t-\rho z_{t-1}) -\beta(D_t-\rho D_{t-1})]^2\}^{-\frac{n}{2}}\{[(\theta-0.9)/0.2] . [1-(\theta-0.9)/0.2)]\}^{48.5}\{(\beta/6)(1-\beta/6)\}^{0.75} .\{(\rho/0.99)(1-\rho/0.99)\}^{1.5628}d\theta.d\beta.d\rho$$

VII. B.P.O. AND POINT PREDICTIONS

(1) B.P.O. Analysis

In each of the four parts of the study, the prior p.d.f.'s and p.m.f.'s are used in conjunction with the appropriate likelihood function to compute the posterior masses for the different models or specifications. The analysis proceeds in the manner described in Section VII of Chapter VII, especially equations VII.VII.2 and VII.VII.3. The actual forms of the functions involved will be apparent from the details in Section VI above, and these are not repeated here.

The values[12] of $p(M_i|y)$ and $p(S_i|y)$ are computed at the end of the basic sample period of forty three

12. Henceforth, for convenience, we abbreviate $p(M_i|y,y_0)$ and $p(S_i|y,y_0)$ by $p(M_i|y)$ and $p(S_i|y)$ respectively.

observations, and also after the addition of each of three further "forecast" observations. That is, $p(M_i|y)$ and $p(S_i|y)$ are updated sequentially, always being based on a sample beginning in 1961(3). For a detailed analysis of the stability of the different models and specifications it would be interesting to evaluate their posterior masses at each observation of the basic sample (after account is taken of degrees of freedom), but the computational cost precluded this possibility.

(2) Point Predictions

Specializing the analysis of Section IV of Chapter II so that y_F denotes just a single "future" observation on the dependent variable, the "predictive p.d.f." is given by equations II.IV.1, II.II.4 and II.IV.2. Clearly, S_i can replace M_i in these equations.

Under the assumption of a p.d. quadratic loss function, point predictions are obtained as the mean of the predictive p.d.f.:

$$\hat{y}_F = \mathcal{E}(y_F|M_i,y) = \int p(y_F|M_i,y)y_F.dy_F \qquad \text{VIII.VII.1}$$

where S_i may replace M_i, and y_0 may also be relevant.

To ease computational expense we do not evaluate[13] interval predictions (which require the computation of the variances of the predictive p.d.f.'s), and we consider only three forecast observations.

13. See also Courville and Geisel (1971), p.20.

Just as it would be interesting to compute the $p(M_i|y)$ or $p(S_i|y)$ for each successive observation (beyond the base required for positive degrees of freedom), so too it would be helpful to compute the full sequence of predictive p.d.f.'s and their moments.

The usefulness of such information is demonstrated well by Geisel's analysis of alternative aggregate consumption expenditure models. However, bearing in mind that our basic relationship, VIII.II.1, is somewhat naive, economically, there seems little to be gained by undertaking an even more detailed analysis.

Finally it should be noted that the forecasts given in the next Section are all conditional upon the observed values of the exogenous variables in the forecast periods. That is, the forecasts are *ex post*.

VIII. RESULTS

(1) Discussion of Tabulated Results

A number of general patterns emerge in all four parts of the study. These are apparent in the tabulated results, and are merely summarized here.

Of particular interest are the posterior probabilities of the various models and specifications, and the sensitivity of results to the use of *a priori* information and to the allowance for serially correlated disturbances.

TABLE VIII.VIII.1

(Part 1 : Diffuse Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta\|y)$	0.959	0.974	0.987	0.999	0.982	1.005	1.021	1.033	0.971	0.989	1.004	0.970
$\mathcal{E}(\beta\|y)$	-0.390	0.869	1.942	2.801	1.578	3.401	4.374	4.959	0.634	2.119	3.270	4.090
$V(\theta\|y)$	2.350	2.520	2.850	3.250	3.130	3.940	4.720	5.290	2.610	3.140	3.760	3.920
$V(\beta\|y)$	2.483	2.637	2.956	3.344	3.264	4.050	4.752	5.197	2.744	3.263	3.857	4.368
$p(M_i\|y)$	0.279	0.558	0.041	0.004	0.000	0.005	0.000	0.000	0.001	0.112	0.000	0.000
$\bar{R}^2$	0.960	0.970	0.950	0.944	0.928	0.947	0.916	0.909	0.941	0.959	0.932	0.884
D.W.	1.015	1.272	0.858	0.762	0.694	0.873	0.642	0.580	0.800	0.981	0.706	0.648

* Multiply $V(\theta|y)$ by 10^{-4}

TABLE VIII.VIII.2

(Part 1 : Diffuse Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0 \mid y)$	0.616	0.519	0.449	0.395	0.402	0.322	0.269	0.231	0.485	0.396	0.335	0.277
$V(W_0 \mid y)$	97.000	72.000	59.000	51.000	52.000	40.000	33.000	26.000	65.000	50.000	42.000	32.000
$\mathcal{E}(W_1 \mid y)$	0.274	0.292	0.287	0.275	0.357	0.302	0.258	0.225	0.360	0.334	0.300	0.257
$V(W_1 \mid y)$	19.000	23.000	24.000	25.000	41.000	35.000	30.000	25.000	36.000	36.000	33.000	27.000
$\mathcal{E}(W_2 \mid y)$	0.068	0.130	0.162	0.176	0.223	0.241	0.226	0.205	0.126	0.198	0.217	0.205
$V(W_2 \mid y)$	1.000	4.000	8.000	10.000	16.000	23.000	23.000	21.000	4.000	13.000	18.000	18.000
$\mathcal{E}(W_3 \mid y)$	-	0.032	0.072	0.099	-	0.141	0.172	0.173	-	0.062	0.118	0.139
$V(W_3 \mid y)$	-	0.000	2.000	3.000	-	8.000	13.000	15.000	-	1.000	5.000	8.000
$\mathcal{E}(W_4 \mid y)$	-	-	0.018	0.044	-	-	0.097	0.128	-	-	0.035	0.072
$V(W_4 \mid y)$	-	-	0.000	1.000	-	-	4.000	8.000	-	-	0.000	2.000
$\mathcal{E}(W_5 \mid y)$	-	-	-	0.011	-	-	-	0.071	-	-	-	0.020
$V(W_5 \mid y)$	-	-	-	0.000	-	-	-	2.000	-	-	-	0.000
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.630	0.925	1.221	1.518

* Multiply all variances by 10^{-6}.

TABLE VIII.VIII.3

(Part 1 : Diffuse Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\bar{M}$
$p(M_i\|y)$	0.279	0.558	0.041	0.004	0.000	0.005	0.000	0.000	0.001	0.112	0.000	0.000	-
$\hat{y}_F$	284.236	284.923	286.273	287.620	282.595	287.168	289.488	291.099	282.224	284.608	287.042	288.930	284.724
% Res.	2.149	3.802	3.478	3.965	2.149	3.802	4.640	5.222	2.014	2.876	3.756	4.438	2.918
1972 (3)													
$p(M_i\|y)$	0.259	0.561	0.033	0.003	0.000	0.005	0.000	0.000	0.006	0.133	0.002	0.000	-
$\hat{y}_F$	277.439	279.977	282.223	284.227	282.096	284.775	287.872	289.861	281.373	283.142	285.302	287.327	279.862
% Res.	1.978	2.911	3.737	4.473	3.690	4.675	5.813	6.544	3.424	4.075	4.868	5.613	2.869
1972 (4)													
$p(M_i\|y)$	0.230	0.639	0.022	0.002	0.000	0.004	0.000	0.000	0.003	0.100	0.000	0.000	-
$\hat{y}_F$	310.389	308.087	306.347	305.257	305.329	302.539	301.427	301.728	305.950	304.430	303.068	302.481	308.177
% Res.	4.419	3.644	3.059	2.692	2.717	1.778	1.404	1.505	2.925	2.414	1.956	1.758	3.675
1973 (1)													
$p(M_i\|y)$	0.276	0.514	0.034	0.003	0.000	0.006	0.000	0.000	0.006	0.161	0.000	0.000	-

TABLE VIII.VIII.4

(Part 1 : Loose Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta\|y)$	0.975	0.984	0.993	1.000	0.989	1.003	1.012	1.020	0.982	0.993	1.003	0.965
$\mathcal{E}(\beta\|y)$	1.513	1.987	2.500	2.910	2.380	3.156	3.452	3.570	1.918	2.608	3.108	3.457
$V(\theta\|y)$	1.020	1.250	1.490	1.650	1.530	1.810	1.970	2.120	1.250	1.580	1.770	1.690
$V(\beta\|y)$	0.802	1.074	1.307	1.425	1.313	1.503	1.520	1.528	1.056	1.374	1.486	1.476
$p(M_i\|y)$	0.361	0.390	0.080	0.009	0.000	0.008	0.000	0.000	0.012	0.140	0.000	0.000

* Multiply $V(\theta|y)$ by 10^{-4}

TABLE VIII.VIII.5

(Part 1 : Loose Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0\|y)$	0.627	0.525	0.451	0.396	0.405	0.321	0.266	0.228	0.491	0.397	0.334	0.256
$V(W_0\|y)$	42.000	36.000	31.000	26.000	26.000	19.000	14.000	11.000	31.000	25.000	20.000	14.000
$\mathcal{E}(W_1\|y)$	0.279	0.295	0.289	0.275	0.360	0.301	0.256	0.222	0.364	0.335	0.300	0.255
$V(W_1\|y)$	8.000	11.000	13.000	12.000	20.000	16.000	13.000	10.000	17.000	18.000	16.000	12.000
$\mathcal{E}(W_2\|y)$	0.070	0.131	0.162	0.176	0.225	0.241	0.224	0.203	0.127	0.199	0.217	0.204
$V(W_2\|y)$	1.000	2.000	4.000	5.000	8.000	10.000	10.000	8.000	2.000	6.000	8.000	8.000
$\mathcal{E}(W_3\|y)$	-	0.033	0.072	0.099	-	0.140	0.170	0.171	-	0.062	0.118	0.138
$V(W_3\|y)$	-	0.000	1.000	2.000	-	4.000	6.000	6.000	-	1.000	2.000	3.000
$\mathcal{E}(W_4\|y)$	-	-	0.018	0.044	-	-	0.096	0.127	-	-	0.035	0.072
$V(W_4\|y)$	-	-	0.000	0.000	-	-	2.000	3.000	-	-	0.000	1.000
$\mathcal{E}(W_5\|y)$	-	-	-	0.011	-	-	-	0.070	-	-	-	0.020
$V(W_5\|y)$	-	-	-	0.000	-	-	-	1.000	-	-	-	0.000
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.630	0.925	1.221	1.518

* Multiply all variances by 10^{-6}.

TABLE VIII.VIII.6

(Part 1 : Loose Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\bar{M}$
$p(M_i \mid y)$	0.361	0.390	0.080	0.009	0.000	0.008	0.000	0.000	0.012	0.140	0.000	0.000	-
$\hat{y}_F$	287.355	286.750	287.187	287.796	283.886	286.769	287.964	288.776	284.295	285.403	286.776	287.981	286.795
% Res.	3.869	3.650	3.808	4.029	2.615	3.657	4.089	4.383	2.763	3.164	3.660	4.095	3.667
1972 (3)													
$p(M_i \mid y)$	0.321	0.382	0.075	0.008	0.000	0.011	0.000	0.000	0.014	0.188	0.000	0.000	-
$\hat{y}_F$	280.656	282.018	283.485	284.909	283.553	284.963	287.218	288.573	283.540	284.235	285.902	287.154	282.188
% Res.	3.161	3.661	4.201	4.724	4.226	4.744	5.573	6.071	4.221	4.476	5.089	5.549	3.724
1972 (4)													
$p(M_i \mid y)$	0.380	0.383	0.061	0.006	0.000	0.010	0.000	0.000	0.010	0.150	0.000	0.000	-
$\hat{y}_F$	314.454	310.799	308.216	306.564	307.420	303.367	301.605	301.422	308.852	306.141	304.022	303.093	311.215
% Res.	5.786	4.557	3.688	3.132	3.420	2.056	1.464	1.402	3.902	2.990	2.277	1.964	4.700
1973 (1)													
$p(M_i \mid y)$	0.185	0.415	0.108	0.013	0.000	0.019	0.000	0.000	0.024	0.235	0.001	0.000	-

TABLE VIII.VIII.7
(Part 1 : Tight Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta\|y)$	0.989	0.993	0.997	1.000	0.996	1.001	1.004	1.006	0.992	0.997	1.001	0.986
$\mathcal{E}(\beta\|y)$	2.499	2.696	2.841	2.876	2.883	3.036	2.869	2.596	2.697	2.924	2.968	4.868
$V(\theta\|y)$	6.300	6.300	6.400	6.600	6.500	6.800	7.100	7.200	6.300	6.500	6.800	5.000
$V(\beta\|y)$	0.821	0.824	0.849	0.882	0.888	0.948	1.011	1.036	0.842	0.882	0.931	0.452
$p(M_i\|y)$	0.195	0.460	0.144	0.018	0.000	0.013	0.000	0.000	0.021	0.147	0.001	0.000

* Multiply $V(\theta|y)$ by 10^{-5}

TABLE VIII.VIII.8

(Part 1 : Tight Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0\|y)$	0.636	0.530	0.453	0.396	0.407	0.320	0.264	0.225	0.496	0.399	0.334	0.282
$V(W_0\|y)$	26.000	18.000	13.000	10.000	11.000	7.000	5.000	4.000	16.000	10.000	8.000	4.000
$\mathcal{E}(W_1\|y)$	0.282	0.298	0.290	0.275	0.362	0.300	0.254	0.219	0.368	0.337	0.299	0.261
$V(W_1\|y)$	5.000	6.000	5.000	5.000	9.000	6.000	5.000	3.000	9.000	7.000	6.000	4.000
$\mathcal{E}(W_2\|y)$	0.071	0.132	0.163	0.176	0.226	0.240	0.222	0.200	0.129	0.200	0.216	0.209
$V(W_2\|y)$	0.000	1.000	2.000	2.000	3.000	4.000	3.000	3.000	1.000	3.000	3.000	2.000
$\mathcal{E}(W_3\|y)$	-	0.033	0.072	0.099	-	0.140	0.169	0.169	-	0.062	0.118	0.141
$V(W_3\|y)$	-	0.000	0.000	1.000	-	1.000	2.000	2.000	-	0.000	1.000	1.000
$\mathcal{E}(W_4\|y)$	-	-	0.018	0.044	-	-	0.095	0.125	-	-	0.035	0.073
$V(W_4\|y)$	-	-	0.000	0.000	-	-	1.000	1.000	-	-	0.000	0.000
$\mathcal{E}(W_5\|y)$	-	-	-	0.011	-	-	-	0.069	-	-	-	0.021
$V(W_5\|y)$	-	-	-	0.000	-	-	-	0.000	-	-	-	0.000
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.630	0.925	1.221	1.518

* Multiply all variances by 10^{-6}.

TABLE VIII.VIII.9
(Part 1 : Tight Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\overline{M}$
$p(M_i\|y)$	0.195	0.460	0.144	0.018	0.000	0.013	0.000	0.000	0.021	0.147	0.001	0.000	-
$\hat{y}_F$	290.235	288.688	288.109	287.725	285.257	286.421	286.237	285.759	286.433	286.257	286.394	292.814	288.450
% Res.	4.910	4.351	4.142	4.003	3.111	3.532	3.465	3.292	3.536	3.472	3.522	5.842	4.265
1972 (3)													
$p(M_i\|y)$	0.140	0.432	0.154	0.022	0.000	0.021	0.000	0.000	0.030	0.200	0.001	0.000	-
$\hat{y}_F$	283.706	284.256	284.853	287.275	285.207	285.120	286.085	286.210	285.901	285.473	285.721	292.429	284.649
% Res.	4.282	4.484	4.703	5.594	4.834	4.802	5.156	5.202	5.089	4.931	5.022	7.488	4.628
1972 (4)													
$p(M_i\|y)$	0.184	0.510	0.169	0.023	0.001	0.014	0.000	0.000	0.025	0.073	0.002	0.000	-
$\hat{y}_F$	308.250	318.205	313.690	310.232	309.748	304.156	301.127	299.717	311.899	308.049	304.812	309.541	314.301
% Res.	3.699	7.048	5.529	4.366	4.203	2.322	1.303	0.829	4.927	3.632	2.543	4.134	5.735
1973 (1)													
$p(M_i\|y)$	0.043	0.356	0.249	0.051	0.002	0.038	0.000	0.000	0.065	0.192	0.005	0.000	-

TABLE VIII.VIII.10

(Part 2 : Diffuse Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta \| y)$	0.815	0.931	0.962	0.978	0.939	0.971	0.975	0.976	0.922	0.959	0.976	0.945
$\mathcal{E}(\beta \| y)$	2.074	2.153	3.144	3.754	2.426	4.551	4.265	4.150	1.909	3.336	4.108	4.493
$\mathcal{E}(\rho \| y)$	0.681	0.653	0.717	0.754	0.706	0.790	0.795	0.822	0.669	0.744	0.777	0.780
$V(\theta \| y)$	3.329	1.026	0.536	0.442	0.939	0.736	0.879	1.053	1.184	0.614	0.567	0.529
$V(\beta \| y)$	6.506	6.047	6.794	6.746	6.506	6.598	7.215	7.530	5.649	6.854	6.799	5.603
$V(\rho \| y)$	9.439	4.771	3.121	2.488	3.572	2.272	2.036	1.748	4.677	2.866	2.285	2.135
$p(M_i \| y)$	0.044	0.249	0.263	0.121	0.008	0.024	0.001	0.000	0.095	0.173	0.018	0.004

* Multiply $V(\theta|y)$ and $V(\rho|y)$ by 10^{-2}

TABLE VIII.VIII.11

(Part 2 : Diffuse Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0\|y)$	0.524	0.496	0.437	0.387	0.384	0.311	0.256	0.218	0.461	0.384	0.326	0.270
$V(W_0\|y)$	137.560	29.190	11.060	6.920	15.710	7.540	6.090	5.260	29.600	9.830	6.300	4.320
$\mathcal{E}(W_1\|y)$	0.233	0.279	0.280	0.269	0.341	0.291	0.246	0.212	0.341	0.324	0.292	0.250
$V(W_1\|y)$	27.170	9.230	4.530	3.340	12.410	6.620	5.610	4.970	16.240	7.000	5.060	3.700
$\mathcal{E}(W_2\|y)$	0.058	0.124	0.157	0.172	0.213	0.233	0.216	0.194	0.120	0.192	0.211	0.200
$V(W_2\|y)$	1.700	1.820	1.430	1.370	4.850	4.240	4.290	4.150	1.990	2.460	2.650	2.370
$\mathcal{E}(W_3\|y)$	-	0.031	0.070	0.097	-	0.136	0.164	0.164	-	0.060	0.115	0.135
$V(W_3\|y)$	-	1.100	0.280	0.430	-	1.440	2.490	2.960	-	0.240	0.780	1.080
$\mathcal{E}(W_4\|y)$	-	-	0.018	0.043	-	-	0.092	0.121	-	-	0.034	0.070
$V(W_4\|y)$	-	-	0.020	0.090	-	-	0.790	1.620	-	-	0.070	0.290
$\mathcal{E}(W_5\|y)$	-	-	-	0.011	-	-	-	0.067	-	-	-	0.020
$V(W_5\|y)$	-	-	-	0.010	-	-	-	0.490	-	-	-	0.020
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.630	0.925	1.221	1.518

* Multiply all variances by 10^{-4}.

TABLE VIII.VIII.12

(Part 2 : Diffuse Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\bar{M}$
$p(M_i\|y)$	0.044	0.249	0.263	0.121	0.008	0.024	0.001	0.000	0.095	0.173	0.018	0.004	-
$\hat{y}_F$	261.522	270.476	270.900	270.055	271.621	270.412	269.524	267.744	271.401	270.905	269.956	256.179	270.245
% Res.	-5.469	-2.232	-2.079	-2.384	-1.818	-2.255	-2.576	-3.220	-1.898	-2.077	-2.420	-7.400	-2.316
1972 (3)													
$p(M_i\|y)$	0.028	0.251	0.317	0.131	0.006	0.018	0.001	0.000	0.085	0.147	0.014	0.002	-
$\hat{y}_F$	263.115	274.286	276.029	276.855	278.079	277.188	278.297	278.662	277.258	277.815	278.012	278.945	275.774
% Res.	-3.287	0.819	1.460	1.764	2.214	1.886	2.294	2.428	1.912	2.116	2.189	2.532	1.366
1972 (4)													
$p(M_i\|y)$	0.034	0.278	0.334	0.129	0.005	0.014	0.000	0.000	0.073	0.119	0.012	0.002	-
$\hat{y}_F$	293.470	302.420	299.922	297.105	298.658	293.256	289.395	287.726	299.148	297.055	293.696	291.791	299.445
% Res.	-1.273	1.738	0.898	-0.050	0.472	-1.345	-2.644	-3.205	0.637	-0.067	-1.197	-1.838	0.737
1973 (1)													
$p(M_i\|y)$	0.022	0.207	0.365	0.160	0.006	0.016	0.000	0.000	0.088	0.120	0.014	0.002	-

TABLE VIII.VIII.13

(Part 2 : Loose Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta\|y)$	0.977	0.984	0.990	0.995	0.985	0.994	0.999	1.002	0.982	0.989	0.995	0.957
$\mathcal{E}(\beta\|y)$	2.147	2.641	3.006	3.187	2.787	3.382	3.327	3.286	2.524	3.052	3.280	3.418
$\mathcal{E}(\rho\|y)$	0.474	0.561	0.613	0.644	0.604	0.663	0.682	0.711	0.573	0.632	0.661	0.676
$V(\theta\|y)$	2.450	3.100	3.740	4.350	3.870	5.140	6.140	7.360	3.220	4.140	4.940	4.340
$V(\beta\|y)$	1.426	1.616	1.662	1.663	1.671	1.665	1.705	1.732	1.590	1.681	1.675	1.643
$V(\rho\|y)$	1.843	1.731	1.574	1.442	1.592	1.351	1.224	1.074	1.678	1.496	1.358	1.220
$p(M_i\|y)$	0.019	0.247	0.306	0.131	0.007	0.024	0.001	0.000	0.098	0.148	0.017	0.002

* Multiply $V(\theta|y)$ by 10^{-4} and $V(\rho|y)$ by 10^{-2}

TABLE VIII.VIII.14

(Part 2 : Loose Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0\|y)$	0.628	0.525	0.450	0.394	0.403	0.318	0.263	0.224	0.491	0.396	0.332	0.274
$V(W_0\|y)$	1.010	0.880	0.770	0.680	0.650	0.530	0.430	0.370	0.800	0.660	0.550	0.350
$\mathcal{E}(W_1\|y)$	0.279	0.295	0.288	0.273	0.358	0.298	0.252	0.218	0.364	0.339	0.297	0.253
$V(W_1\|y)$	0.200	0.280	0.320	0.330	0.510	0.460	0.390	0.350	0.440	0.470	0.440	0.300
$\mathcal{E}(W_2\|y)$	0.070	0.131	0.162	0.175	0.224	0.239	0.221	0.199	0.127	0.198	0.215	0.203
$V(W_2\|y)$	0.010	0.060	0.100	0.130	0.200	0.300	0.300	0.290	0.050	0.170	0.230	0.190
$\mathcal{E}(W_3\|y)$	-	0.033	0.072	0.098	-	0.139	0.168	0.168	-	0.062	0.117	0.137
$V(W_3\|y)$	-	0.000	0.020	0.040	-	0.100	0.170	0.210	-	0.020	0.070	0.090
$\mathcal{E}(W_4\|y)$	-	-	0.018	0.044	-	-	0.095	0.124	-	-	0.034	0.071
$V(W_4\|y)$	-	-	0.000	0.010	-	-	0.060	0.110	-	-	0.010	0.020
$\mathcal{E}(W_5\|y)$	-	-	-	0.011	-	-	-	0.068	-	-	-	0.020
$V(W_5\|y)$	-	-	-	0.000	-	-	-	0.030	-	-	-	0.000
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.693	0.925	1.221	1.518

* Multiply all variances by 10^{-4}.

TABLE VIII.VIII.15

(Part 2 : Loose Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\bar{M}$
$p(M_i\|y)$	0.019	0.247	0.306	0.131	0.007	0.024	0.001	0.000	0.098	0.148	0.017	0.002	-
$\hat{y}_F$	281.480	277.637	275.320	273.625	271.415	270.659	269.845	267.980	274.671	272.144	271.030	270.863	275.027
% Res.	1.746	0.356	-0.481	-1.094	-1.893	-2.166	-2.460	-3.134	-0.716	-1.629	-2.032	-2.092	-0.587
1972 (3)													
$p(M_i\|y)$	0.017	0.286	0.352	0.138	0.006	0.022	0.000	0.000	0.010	0.168	0.001	0.000	-
$\hat{y}_F$	293.061	279.772	278.880	279.137	281.276	279.684	280.852	281.335	282.498	280.419	280.141	280.039	279.740
% Res.	7.720	2.836	2.508	2.602	3.389	2.804	3.233	3.410	3.838	3.074	2.971	2.934	2.824
1972 (4)													
$p(M_i\|y)$	0.019	0.296	0.335	0.121	0.004	0.016	0.000	0.000	0.072	0.123	0.010	0.001	-
$\hat{y}_F$	310.747	305.747	301.781	298.791	300.918	295.150	291.569	289.639	302.753	298.974	295.467	293.416	301.402
% Res	4.539	2.857	1.523	0.517	1.233	-0.708	-1.912	-2.562	1.850	0.599	-0.601	-1.291	1.395
1973 (1)													
$p(M_i'\|y)$	0.005	0.206	0.380	0.160	0.006	0.009	0.000	0.000	0.094	0.125	0.014	0.001	-

TABLE VIII.VIII.16

(Part 2 : Tight Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(\theta\|y)$	0.992	0.995	0.997	0.999	0.997	0.999	1.000	1.001	0.996	0.998	0.999	0.992
$\mathcal{E}(\beta\|y)$	2.913	3.114	3.258	3.291	3.241	3.498	3.296	3.160	3.163	3.364	3.384	4.275
$\mathcal{E}(\rho\|y)$	0.477	0.549	0.594	0.624	0.585	0.641	0.663	0.696	0.561	0.612	0.640	0.703
$V(\theta\|y)$	7.100	8.500	9.600	10.100	6.900	6.800	7.100	7.500	7.300	6.700	6.700	9.800
$V(\beta\|y)$	1.192	1.271	1.328	1.375	1.293	1.333	1.458	1.553	1.260	1.279	1.341	1.113
$V(\rho\|y)$	1.636	1.556	1.448	1.344	1.462	1.262	1.139	0.990	1.508	1.384	1.271	1.248
$p(M_i\|y)$	0.012	0.228	0.330	0.145	0.007	0.024	0.001	0.000	0.108	0.126	0.019	0.000

* Multiply $V(\theta|y)$ by 10^{-5}, and $V(\rho|y)$ by 10^{-2}.

TABLE VIII.VIII.17

(Part 2 : Tight Prior)*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
$\mathcal{E}(W_0\|y)$	0.638	0.530	0.453	0.395	0.408	0.320	0.263	0.224	0.498	0.399	0.333	0.284
$V(W_0\|y)$	2.900	2.400	2.000	1.600	1.200	0.700	0.500	0.400	1.800	1.100	0.700	0.800
$\mathcal{E}(W_1\|y)$	0.283	0.298	0.290	0.274	0.363	0.300	0.253	0.218	0.369	0.337	0.298	0.262
$V(W_1\|y)$	0.600	0.800	0.800	0.800	0.900	0.600	0.500	0.400	1.000	0.800	0.600	0.700
$\mathcal{E}(W_2\|y)$	0.071	0.133	0.163	0.176	0.227	0.240	0.221	0.199	0.129	0.200	0.216	0.210
$V(W_2\|y)$	0.000	0.200	0.300	0.300	0.400	0.400	0.300	0.300	0.100	0.300	0.300	0.400
$\mathcal{E}(W_3\|y)$	-	0.033	0.072	0.099	-	0.140	0.168	0.168	-	0.062	0.117	0.142
$V(W_3\|y)$	-	0.000	0.100	0.100	-	0.100	0.200	0.200	-	0.000	0.100	0.200
$\mathcal{E}(W_4\|y)$	-	-	0.018	0.044	-	-	0.095	0.124	-	-	0.035	0.074
$V(W_4\|y)$	-	-	0.000	0.000	-	-	0.100	0.100	-	-	0.000	0.100
$\mathcal{E}(W_5\|y)$	-	-	-	0.011	-	-	-	0.068	-	-	-	0.021
$V(W_5\|y)$	-	-	-	0.000	-	-	-	0.000	-	-	-	0.000
A.L.	0.429	0.667	0.909	1.154	0.818	1.200	1.579	1.956	0.630	0.925	1.221	1.518

* Multiply all variances by 10^{-5}.

TABLE VIII.VIII.18

(Part 2 : Tight Prior)

1972 (2)	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$\overline{M}$
$p(M_i \mid y)$	0.012	0.228	0.330	0.145	0.007	0.024	0.001	0.000	0.108	0.126	0.019	0.000	-
$\hat{y}_F$	283.467	279.111	276.452	274.462	273.051	271.634	270.412	271.457	276.498	273.454	271.927	283.843	276.250
% Res.	2.464	0.889	-0.072	-0.791	-1.301	-1.814	-2.255	-1.878	-0.055	-1.156	-1.708	2.600	-0.145
1972 (3)													
$p(M_i \mid y)$	0.008	0.227	0.353	0.146	0.006	0.020	0.000	0.000	0.096	0.129	0.015	0.000	-
$\hat{y}_F$	297.216	282.567	280.595	279.910	283.309	280.193	280.347	279.684	285.400	281.920	280.586	288.321	281.716
% Res.	9.248	3.863	3.138	2.886	4.136	2.991	3.047	2.804	4.905	3.625	3.135	5.978	3.550
1972 (4)													
$p(M_i \mid y)$	0.009	0.254	0.378	0.149	0.005	0.016	0.001	0.000	0.082	0.094	0.013	0.000	-
$\hat{y}_F$	313.624	307.667	303.042	299.587	302.487	295.849	291.688	289.575	304.835	300.194	296.172	296.513	303.761
% Res.	5.507	3.503	1.947	0.785	1.760	-0.473	-1.872	-2.577	2.550	0.989	-0.364	-0.249	2.189
1973 (1)													
$p(M_i \mid y)$	0.002	0.138	0.414	0.210	0.007	0.017	0.001	0.000	0.082	0.110	0.019	0.000	-

TABLE VIII.VIII.19

Part 3*

	Diffuse			Loose			Tight		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
$\mathcal{E}(\theta \mid y)$	0.970	1.005	0.989	0.982	1.003	0.992	0.993	1.001	0.996
$\mathcal{E}(\beta \mid y)$	0.529	3.401	2.106	1.844	3.156	2.554	2.679	3.036	2.896
$\mathcal{E}(L \mid y)$	3.739	4.000	3.991	3.691	4.000	3.921	3.982	4.000	3.882
$V(\theta \mid y)$	26.133	39.400	0.130	34.501	1.810	42.528	11.865	6.800	86.933
$V(\beta \mid y)$	3.063	4.050	3.277	1.091	1.503	1.381	0.840	0.948	0.883
$V(L \mid y)$	0.313	0.000	0.009	0.470	0.000	0.073	0.503	0.000	0.116
$p(S_i \mid y)$	0.882	0.005	0.113	0.839	0.009	0.152	0.817	0.013	0.169

* Multiply $V(\theta|y)$ by 10^{-5}.

TABLE VIII.VIII.20

Part 3*

	Diffuse			Loose			Tight		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
$\mathcal{E}(W_0\|y)$	0.546	0.322	0.397	0.561	0.321	0.404	0.539	0.320	0.411
$V(W_0\|y)$	248.634	207.408	490.000	366.834	190.000	100.834	382.283	70.000	79.646
$\mathcal{E}(W_1\|y)$	0.286	0.302	0.334	0.288	0.301	0.337	0.292	0.300	0.341
$V(W_1\|y)$	8.877	3.500	19.568	21.339	1.600	27.429	21.118	60.000	13.207
$\mathcal{E}(W_2\|y)$	0.112	0.241	0.197	0.108	0.241	0.193	0.124	0.240	0.191
$V(W_2\|y)$	97.423	2.300	20.144	12.579	1.000	50.462	100.383	40.000	65.853
$\mathcal{E}(W_3\|y)$	0.024	0.141	0.061	0.023	0.140	0.057	0.033	0.140	0.055
$V(W_3\|y)$	35.738	1.300	8.997	57.677	0.400	29.245	65.382	10.000	40.104
$\mathcal{E}(W_4\|y)$	0.001	0.000	0.000	0.002	0.000	0.000	0.004	0.000	0.000
$V(W_4\|y)$	2.284	0.400	0.000	4.766	0.000	0.000	8.377	0.000	0.728
$\mathcal{E}(W_5\|y)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$V(W_5\|y)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A.L.	0.790	1.388	1.074	0.790	1.388	1.074	0.790	1.388	1.074

* multiply all variances by 10^{-5}.

TABLE VIII.VIII.21

Part 3

	Diffuse				Loose				Tight			
1972(2)	S_1	S_2	S_3	$\bar{S}$	S_1	S_2	S_3	$\bar{S}$	S_1	S_2	S_3	$\bar{S}$
$p(S_i\|y)$	0.882	0.005	0.113	-	0.839	0.009	0.152	-	0.817	0.014	0.169	-
$\hat{y}_F$	283.476	287.168	284.587	283.620	287.405	286.769	285.316	287.082	288.934	286.421	286.280	288.450
% Res.	2.266	3.597	2.667	2.317	3.683	3.454	2.929	3.566	4.235	3.328	3.277	4.060
1972(3)												
$p(S_i\|y)$	0.855	0.006	0.139	-	0.786	0.012	0.202	-	0.747	0.021	0.232	-
$\hat{y}_F$	279.637	284.775	287.171	280.715	281.631	284.963	284.187	282.187	284.745	285.120	284.299	284.649
% Res.	0.053	1.891	2.748	0.438	0.766	1.958	1.681	0.965	1.880	2.014	1.721	1.846
1972(4)												
$p(S_i\|y)$	0.893	0.004	0.103	-	0.830	0.010	0.160	-	0.886	0.015	0.099	-
$\hat{y}_F$	308.631	302.539	304.474	308.178	312.252	303.367	306.310	311.212	315.069	304.156	312.067	314.608
% Res.	3.085	1.050	1.696	2.933	4.294	1.326	2.309	3.947	5.235	1.590	4.232	5.081
1973(1)												
$p(S_i\|y)$	0.827	0.006	0.167	-	0.721	0.019	0.260	-	0.699	0.040	0.261	-

TABLE VIII.VIII.22

Part 4*

	Diffuse			Loose			Tight		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
$\mathcal{E}(\theta\|y)$	0.944	0.963	0.948	0.988	0.992	0.987	0.997	0.999	0.997
$\mathcal{E}(\beta\|y)$	2.819	4.027	2.932	2.888	3.250	2.874	3.213	3.435	3.280
$\mathcal{E}(\rho\|y)$	0.698	0.770	0.722	0.597	0.651	0.612	0.584	0.629	0.592
$\mathcal{E}(L\|y)$	4.681	3.788	3.762	4.781	3.812	3.709	4.850	3.812	3.648
$V(\theta\|y)$	1.010	0.879	0.782	0.131	0.088	0.047	0.028	0.074	0.051
$V(\beta\|y)$	6.888	7.419	7.001	1.700	1.729	1.723	1.324	1.343	1.284
$V(\rho\|y)$	4.135	2.678	3.554	1.721	1.421	1.695	1.533	1.414	1.546
$V(L\|y)$	0.705	0.228	0.388	0.598	0.215	0.332	0.566	0.215	0.378
$p(S_i\|y)$	0.677	0.033	0.290	0.703	0.032	0.265	0.715	0.032	0.253

* Multiply $V(\theta|y)$ and $V(\rho|y)$ by 10^{-2}.

TABLE VIII.VIII.23

Part 4*

	Diffuse			Loose			Tight		
	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
$\mathcal{E}(W_0 \mid y)$	0.455	0.327	0.404	0.471	0.335	0.426	0.469	0.337	0.436
$V(W_0 \mid y)$	4.781	2.094	3.571	2.810	1.362	2.967	2.818	1.814	2.398
$\mathcal{E}(W_1 \mid y)$	0.275	0.302	0.327	0.287	0.310	0.345	0.289	0.312	0.348
$V(W_1 \mid y)$	0.675	1.205	0.927	0.333	0.570	0.343	0.190	0.987	0.248
$\mathcal{E}(W_2 \mid y)$	0.141	0.228	0.170	0.151	0.235	0.173	0.155	0.237	0.171
$V(W_2 \mid y)$	1.014	0.350	1.356	0.500	0.146	1.220	0.247	0.167	1.209
$\mathcal{E}(W_3 \mid y)$	0.056	0.104	0.045	0.061	0.110	0.043	0.064	0.110	0.040
$V(W_3 \mid y)$	0.862	3.563	1.215	0.709	3.281	0.736	0.632	3.280	0.342
$\mathcal{E}(W_4 \mid y)$	0.015	0.003	0.003	0.016	0.003	0.003	0.017	0.003	0.003
$V(W_4 \mid y)$	0.234	0.250	0.131	0.246	0.273	0.103	0.253	0.029	0.101
$\mathcal{E}(W_5 \mid y)$	0.002	0.000	0.000	0.002	0.000	0.000	0.002	0.000	0.000
$V(W_5 \mid y)$	0.018	0.000	0.000	0.019	0.000	0.000	0.021	0.000	0.000
A.L.	0.790	1.388	1.074	0.790	1.388	1.074	0.790	1.388	1.074

* Multiply all variances by 10^{-3}.

TABLE VIII.VIII.24

Part 4*

1972 (2)	Diffuse S_1	Diffuse S_2	Diffuse S_3	Diffuse $\bar{S}$	Loose S_1	Loose S_2	Loose S_3	Loose $\bar{S}$	Tight S_1	Tight S_2	Tight S_3	Tight $\bar{S}$
$p(S_i\|y)$	0.677	0.033	0.290	-	0.703	0.031	0.266	-	0.715	0.032	0.253	-
$\hat{y}_F$	269.984	270.678	270.805	290.245	275.985	270.799	272.997	275.029	277.014	271.906	274.639	276.250
% Res.	-2.602	-2.351	-2.306	-2.508	-0.437	-2.308	-1.538	-0.782	-0.066	-1.908	-0.922	-0.341
1972 (3)												
$p(S_i\|y)$	0.727	0.024	0.249	-	0.794	0.027	0.179	-	0.734	0.026	0.240	-
$\hat{y}_F$	275.079	289.006	276.529	275.774	279.198	280.025	280.534	279.459	281.250	280.912	283.229	281.716
% Res.	-1.578	3.405	-1.059	-1.330	-0.105	0.191	0.374	-0.011	0.630	0.509	1.338	0.796
1972 (4)												
$p(S_i\|y)$	0.774	0.020	0.206	-	0.772	0.020	0.208	-	0.790	0.021	0.189	-
$\hat{y}_F$	300.454	294.678	297.550	299.740	302.663	296.304	297.212	301.402	303.998	297.168	301.931	303.464
% Res.	0.353	-1.576	-0.617	0.115	1.091	-1.033	-0.729	0.670	1.537	-0.744	0.847	1.359
1973 (1)												
$p(S_i\|y)$	0.754	0.022	0.224	-	0.751	0.015	0.234	-	0.764	0.025	0.211	-

The tables of results are categorized at three levels: first, according to the part of the study; secondly, according to the form of the prior p.d.f.'s (i.e. diffuse, loose, or tight); and thirdly, according to the type of information being presented.

The latter information takes three forms: first, moments of the posterior p.d.f.'s for the parameters, and posterior probabilities for the models or specifications for the basic sample; secondly, moments of the posterior p.d.f.'s for the individual w_k's in VIII.II.1 obtained from the posterior p.d.f.'s for θ; and thirdly, posterior probabilities for the models or specifications, and point forecasts for three additional periods. The notation used in the tables is as follows:

(i) $\mathcal{E}(\phi|y)$ is the mean of the marginal posterior p.d.f. for any parameter ϕ. Conditionality on M_i or S_i (and on y_0 where appropriate) is implicit here and elsewhere in this notation.

$V(\phi|y)$ is the variance of the marginal posterior p.d.f. for any parameter ϕ.

$p(M_i|y)$ and $p(S_i|y)$ are the posterior probabilities of the ith. model and the ith. specification respectively, as at the start of the quarter shown.

(ii) w_k is the kth. weight in the lag distribution for equation VIII.II.1.

A.L. is the "average lag length",

$(\sum_{k=0}^{L} kw_k)/(\sum_{k=0}^{L} w_k)$. See Giles (1974a).

(iii) $\hat{y}_F$ is the mean of the predictive p.d.f. for y in the forecast period, F.

% Res. is the percentage forecast residual, $100(\hat{y}_F - y_F)/y_F$.

$\overline{M}$ is an average result across the twelve different models. the weights being the $p(M_i|y)$.

$\overline{S}$ is an average result across the three different specifications. the weights being the $p(S_i|y)$.

In all four parts of the study, the results obtained when no a priori information is used (i.e. "diffuse" prior p.d.f.'s are used for the parameters) do not contradict our general a priori feelings, in all but one instance.[14] Further, as additional prior information is introduced via the prior p.d.f.'s, the posterior means move closer to the prior means, and the posterior variances become smaller, as expected.

Marginal Bayes confidence intervals[15] may be constructed for the various parameters. For example, considering only diffuse priors, the value $\theta=1.0$ lies in a 95% Bayes confidence interval in every case, except for models M_1 and M_9 in Table VIII.VIII.1. Further the value $\rho=0.0$ never lies in a 95% Bayes confidence interval in Part (2) of the study, supporting the hypothesis of positive first-order serial correlation.

The residuals computed from the estimated relation-

---

14. The exception is that $\mathcal{E}(\beta|y)$ is negative for M_1 in Table VIII.VIII.1.
15. See Box and Tiao (1973), pp.120-122, for a discussion of such intervals, and comparisons with joint confidence regions.

ships show distinctive patterns, suggesting serial correlation, or model mis-specification. The latter may arise from the omission of relevant explanatory variables, inappropriate lag shapes or lengths, or inappropriate seasonal adjustment. The residual pattern is less marked in Parts (2) and (4) of the study where first-order serial correlation is accounted for explicitly, but the fact that it is apparent even there, suggests that there is still some model mis-specification.

These residual "runs" are reflected in two ways in the tabulated results: first, the values of the Durbin Watson (D.W.) statistic are very low[16] in Table VIII.VIII.1; and secondly, the forecast errors in Tables VIII.VIII.3, and VIII.VIII.6, etc., often exhibit either persistent over-estimation or under-estimation.

The posterior probabilities of the different models are generally insensitive to the choice of prior p.d.f.'s for the parameters. Further, if models are ranked by posterior probability, this ranking is quite insensitive to extensions of the sample period, within a given part of the study.

When no allowance is made for serial correlation, a maximum lag length of L=4 gives rise to models which rank highly in terms of posterior probability. When serial correlation is accounted for, L=5 is a "more probable" maximum lag length. Marginal posterior probabilities for L can be derived from the tabulated data.

16. Conventional tests for the presence of serial correlation are based on O.L.S. parameter estimates. Thus, such tests are inappropriate in this study, except in Table VIII.VIII.1.

The posterior probabilities discriminate markedly among the different models,[17] and models based on values of L other than 4 or 5 are heavily discounted by the sample evidence.

In Parts (3) and (4) of the study the model space collapses to the three elements S_1, S_2 and S_3. The values of $\mathscr{E}(L|y)$ computed here are in accordance with the above statements concerning the maximum lag length. Further, the posterior probabilities favour S_1 rather strongly. Thus, as far as the impact of I on y is concerned, the decay path favoured by the data is one exhibiting rapid initial decay and slow final decay over a period of four to five quarters.

To this extent, the results provide information concerning the underlying economic relationships. Of course, further information is implicitly revealed in the moments of the posterior p.d.f.'s for the various parameters, and especially in those for the w_k's.

(2) <u>Diagrammatic Results</u>

Corresponding to the tabulated results, a number of diagrams are presented to illustrate the shapes of the posterior p.d.f.'s and p.m.f.'s for the various unknown parameters.

Those shown are for the model or specification with the highest posterior probability in that part of the study in question. In each case the posterior p.d.f.'s and

---

17. Compare this with the lack of discrimination afforded by $\bar{R}^2$ in Table VIII.VIII.1.

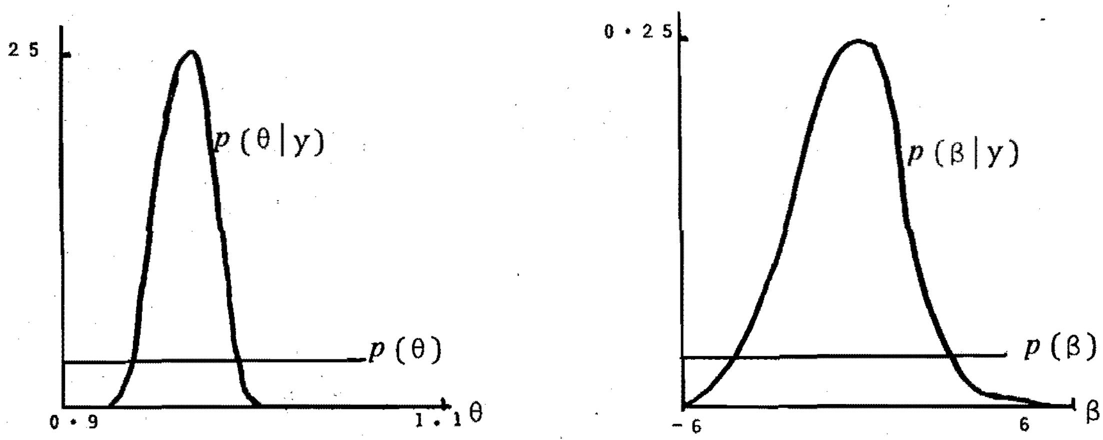


Figure VIII.VIII.1: Part 1, M_2 (Diffuse prior)

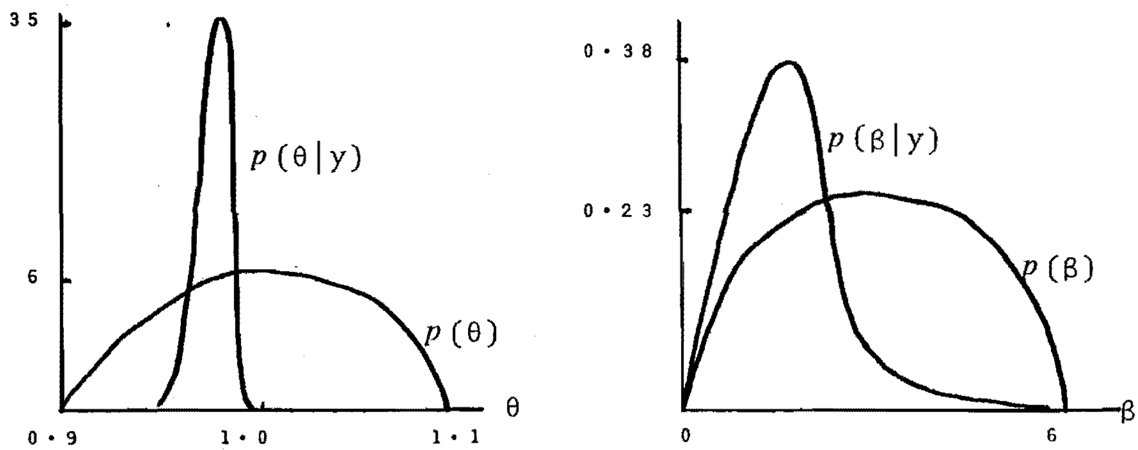


Figure VIII.VIII.2: Part 1, M_2 (Loose prior)

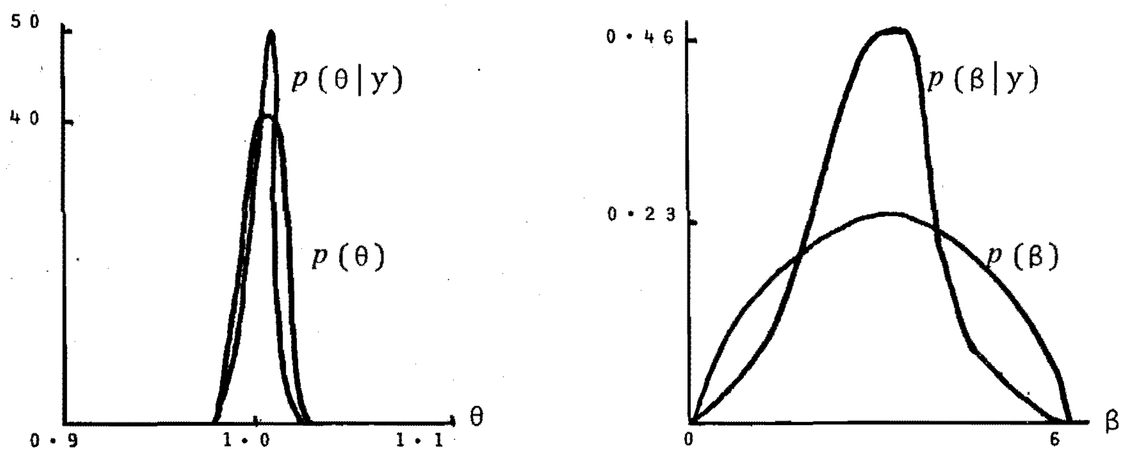


Figure VIII.VIII.3: Part 1, M_2 (Tight Prior)

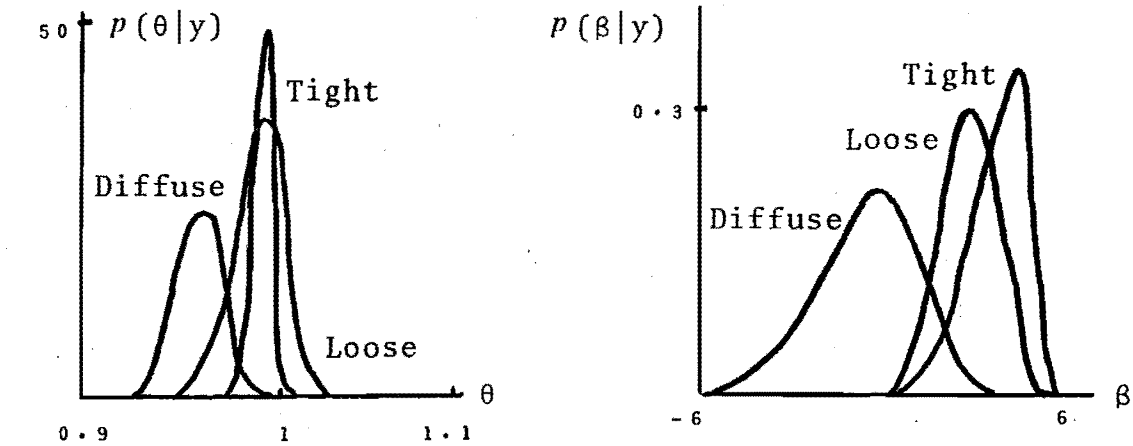


Figure VIII.VIII.4: Part 1, M_2 (posteriors)

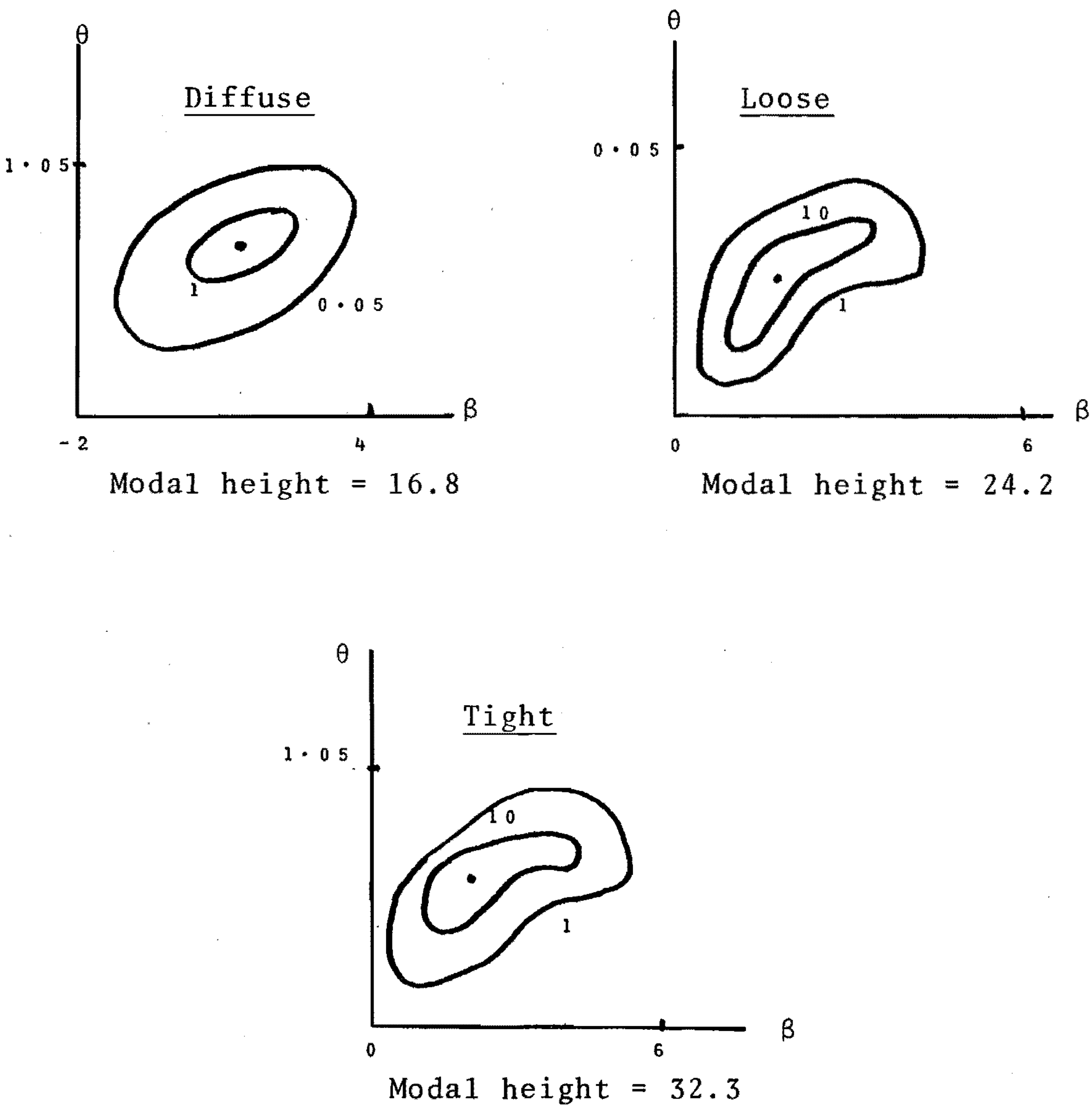


Figure VIII.VIII.5: Part 1, M_2 (Joint posteriors)

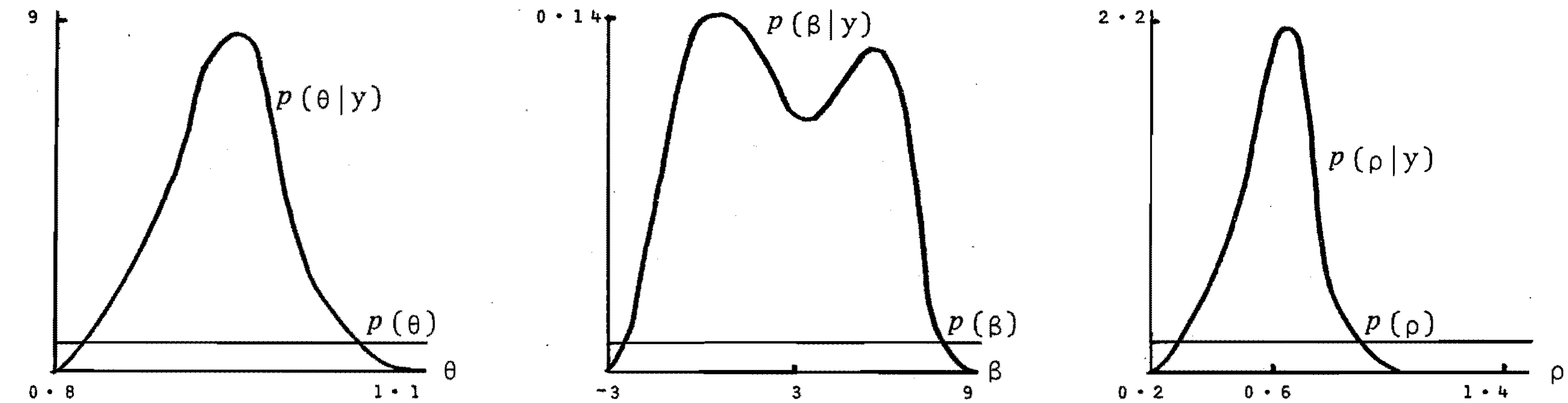


Figure VIII.VIII.6: Part 2, M_3 (Diffuse prior)

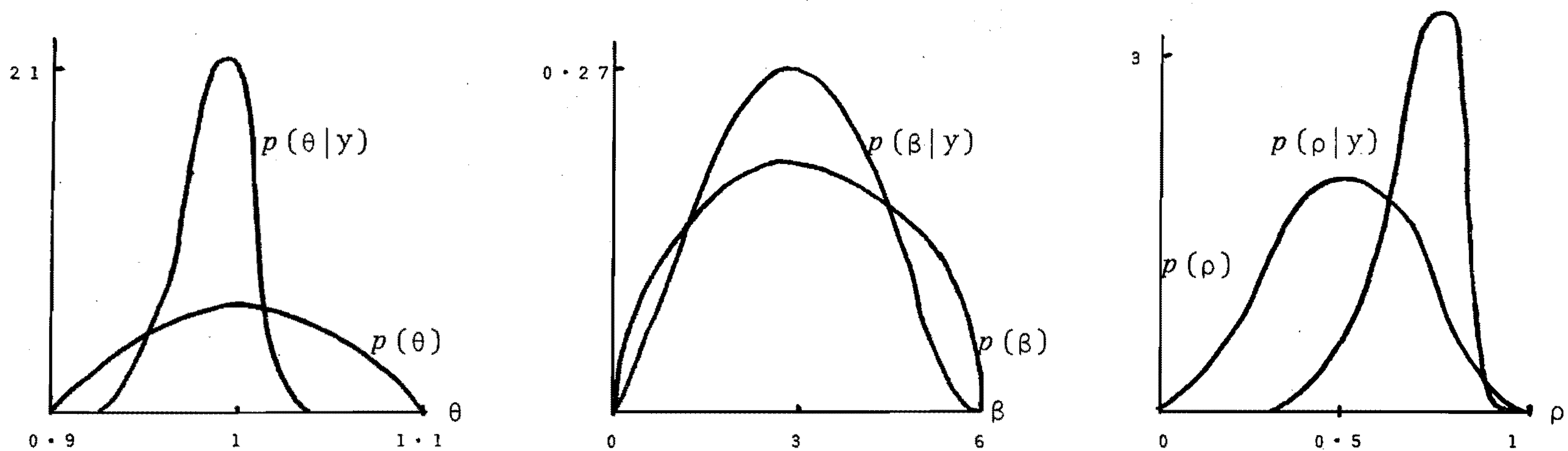


Figure VIII.VIII.7: Part 2, M_3 (Loose prior)

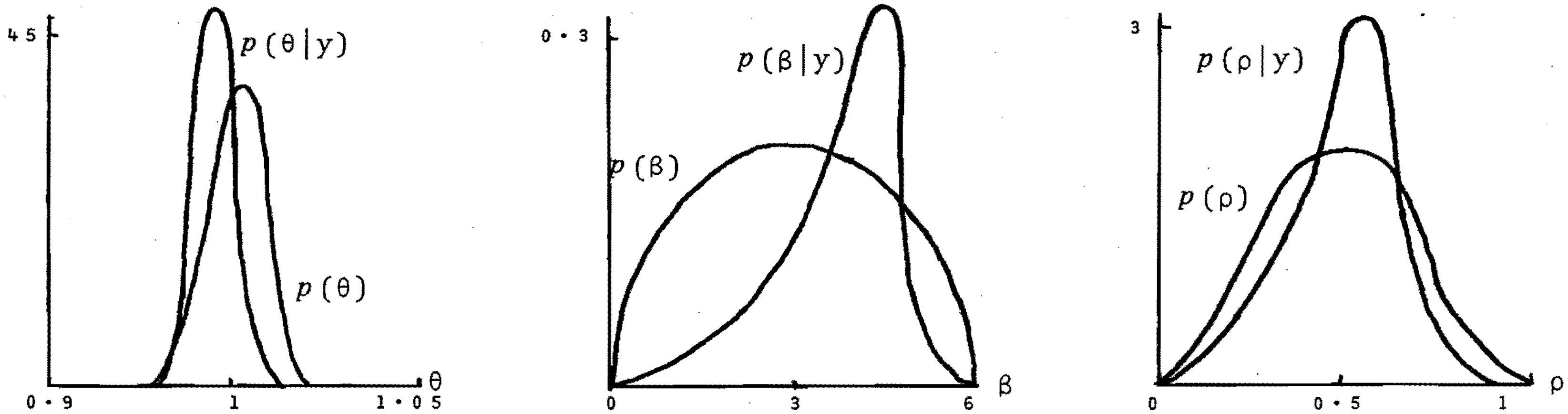


Figure VIII.VIII.8: Part 2, M_3 (Tight prior)

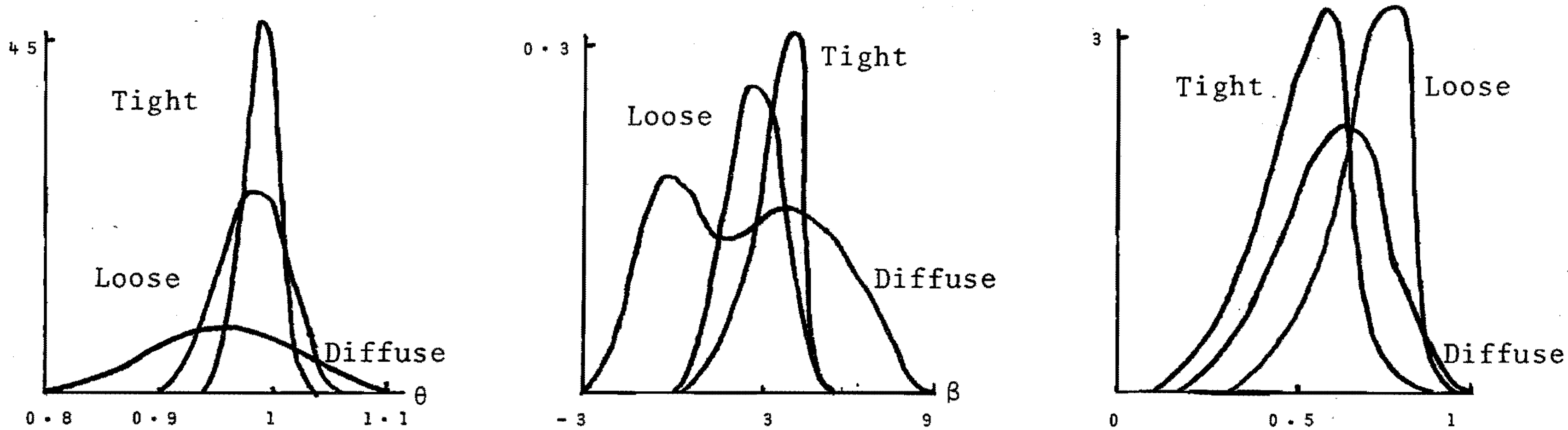


Figure VIII.VIII.9: Part 2, M_3 (Comparative posteriors)

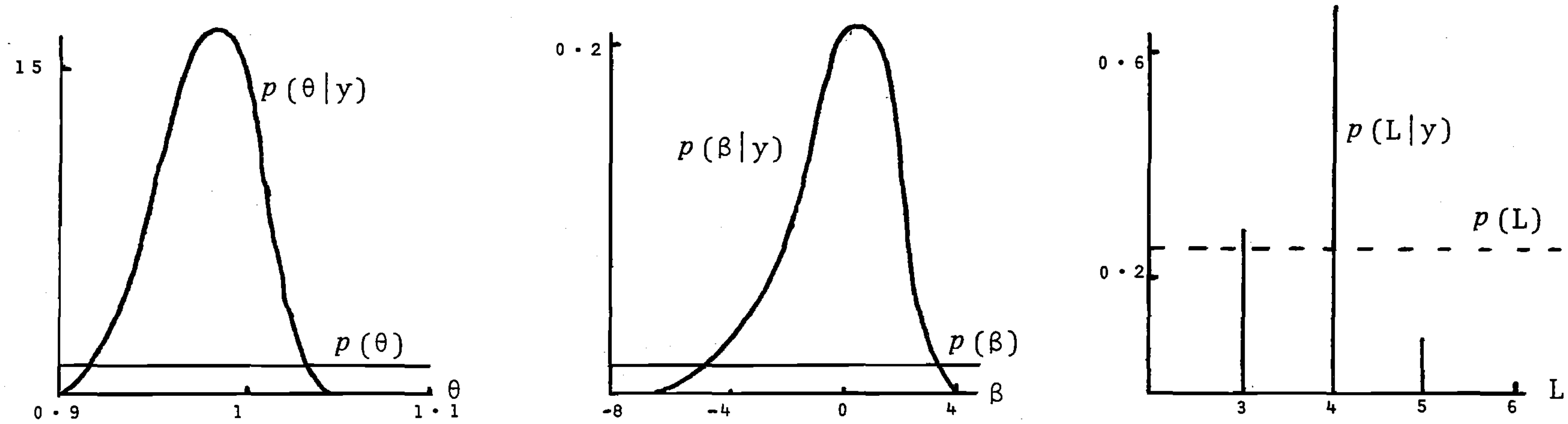


Figure VIII.VIII.10: Part 3, S_1 (Diffuse prior)

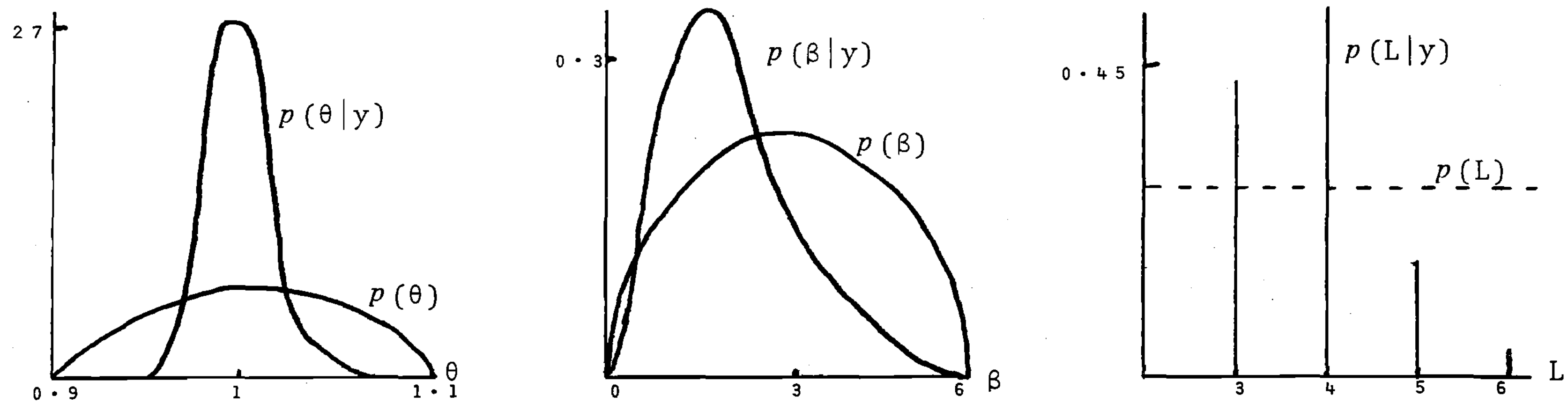


Figure VIII.VIII.11: Part 3, S_1 (Loose prior)

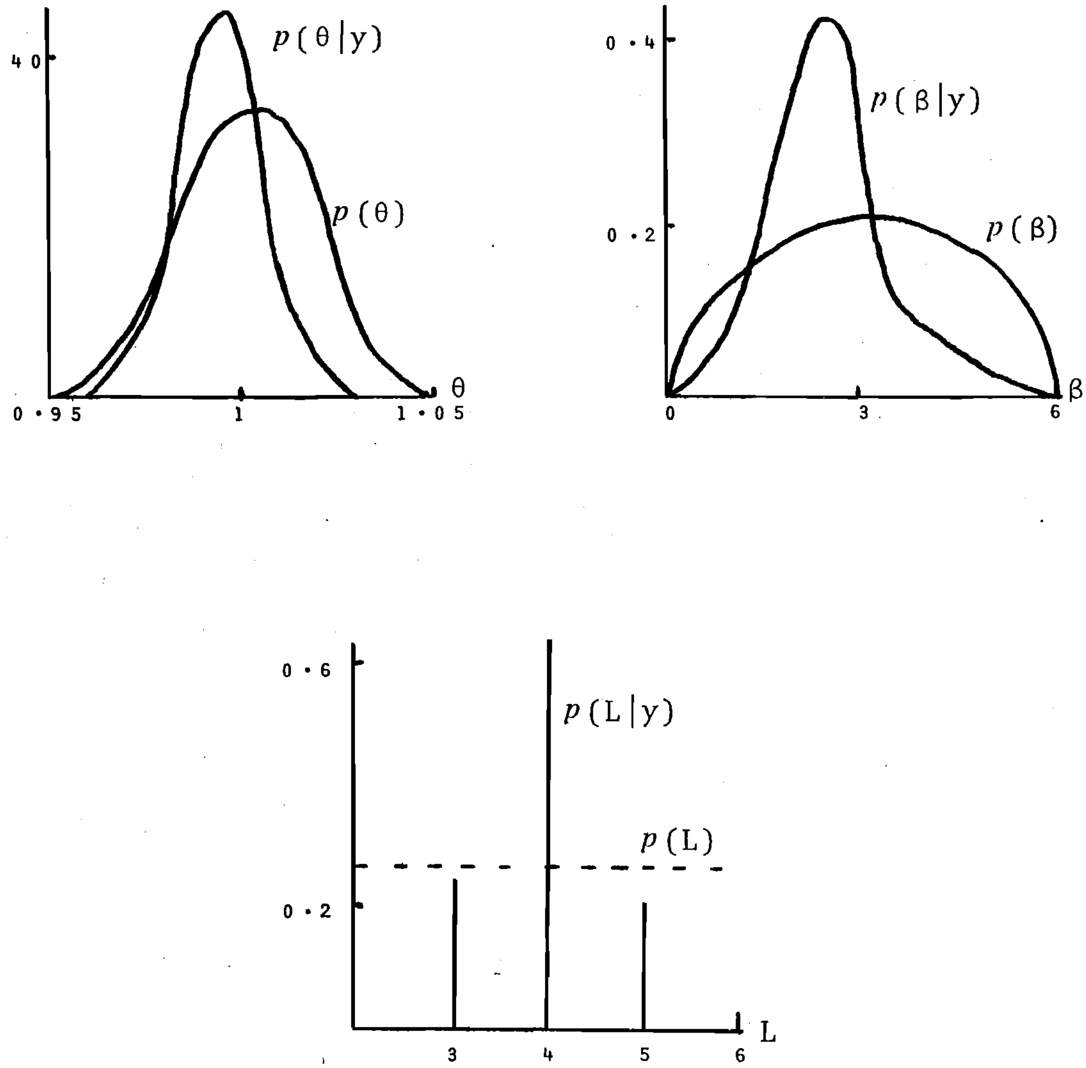


Figure VIII.VIII.12: Part 3, S_1 (Tight prior)

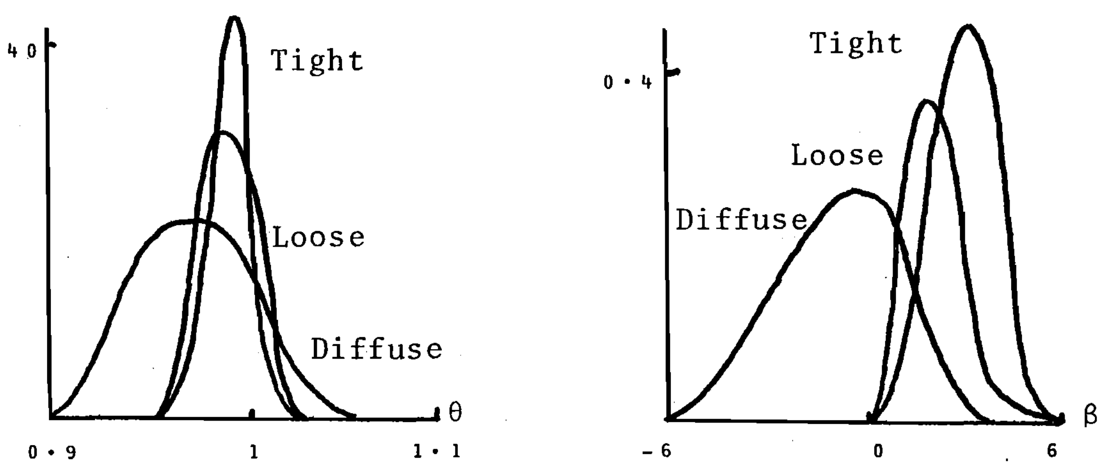


Figure VIII.VIII.13: Part 3, S_1 (Comparative posteriors)

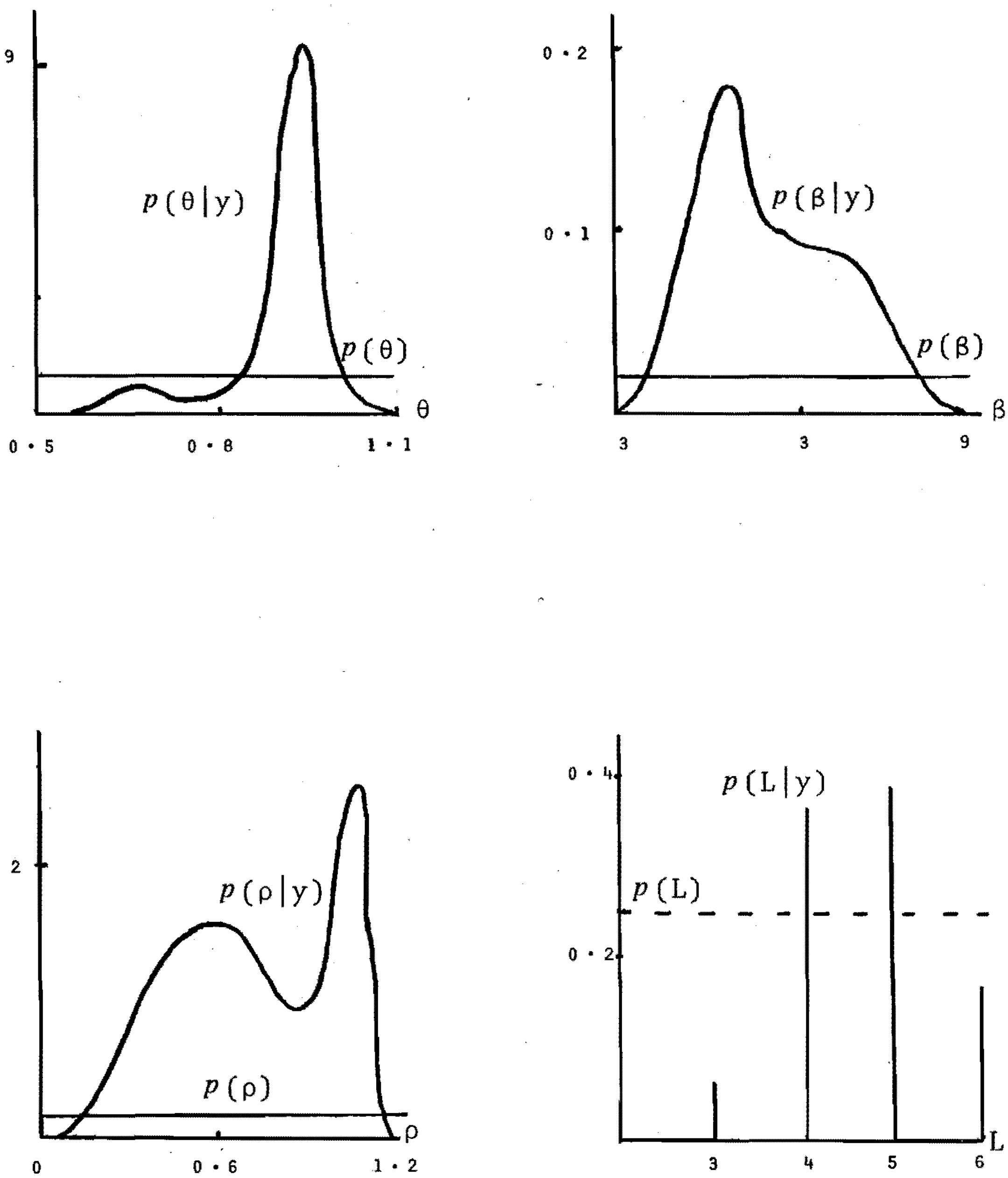


Figure VIII.VIII.14: Part 4, S_1 (Diffuse prior)

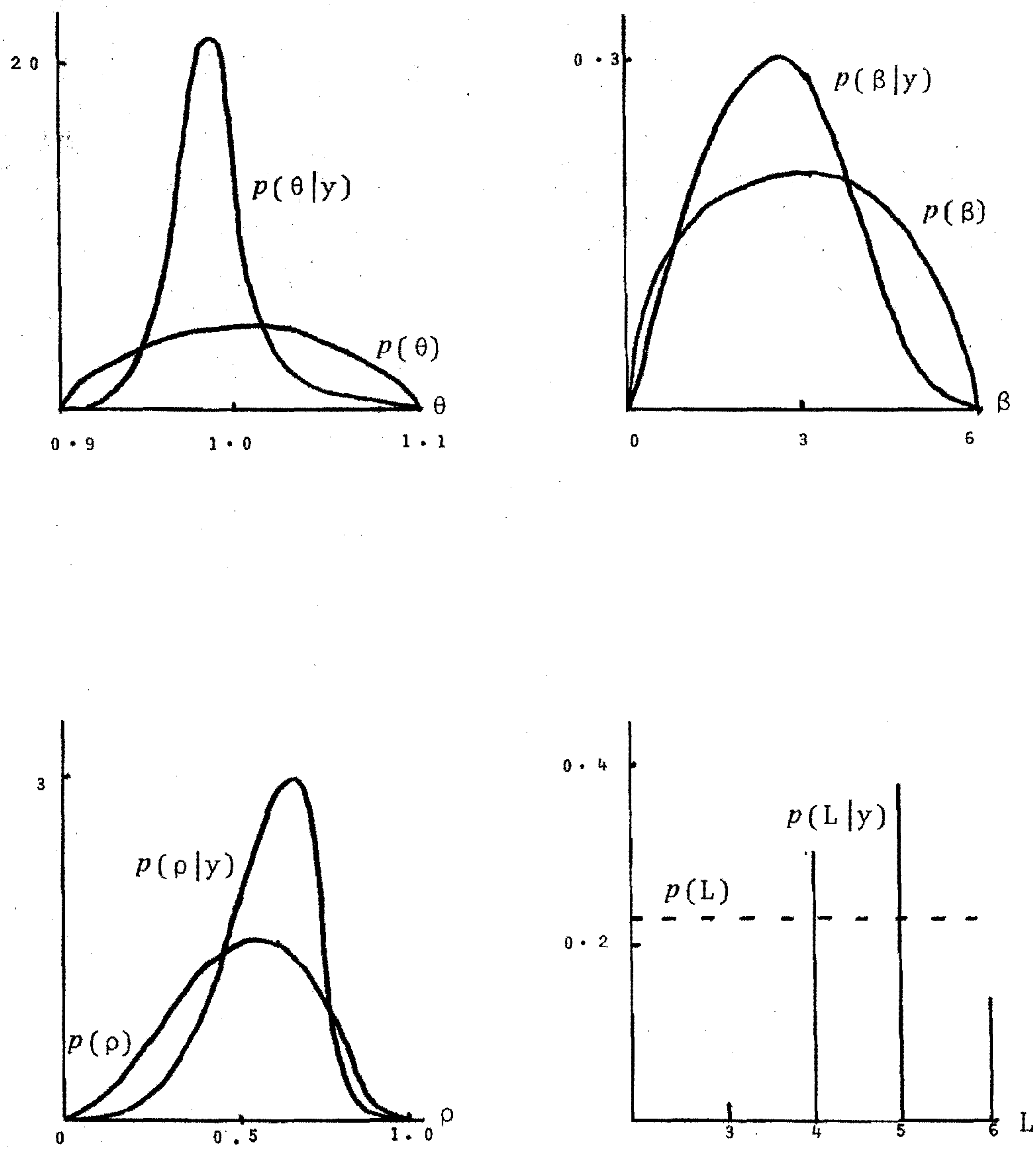


Figure VIII.VIII.15: Part 4, S_1 (Loose prior)

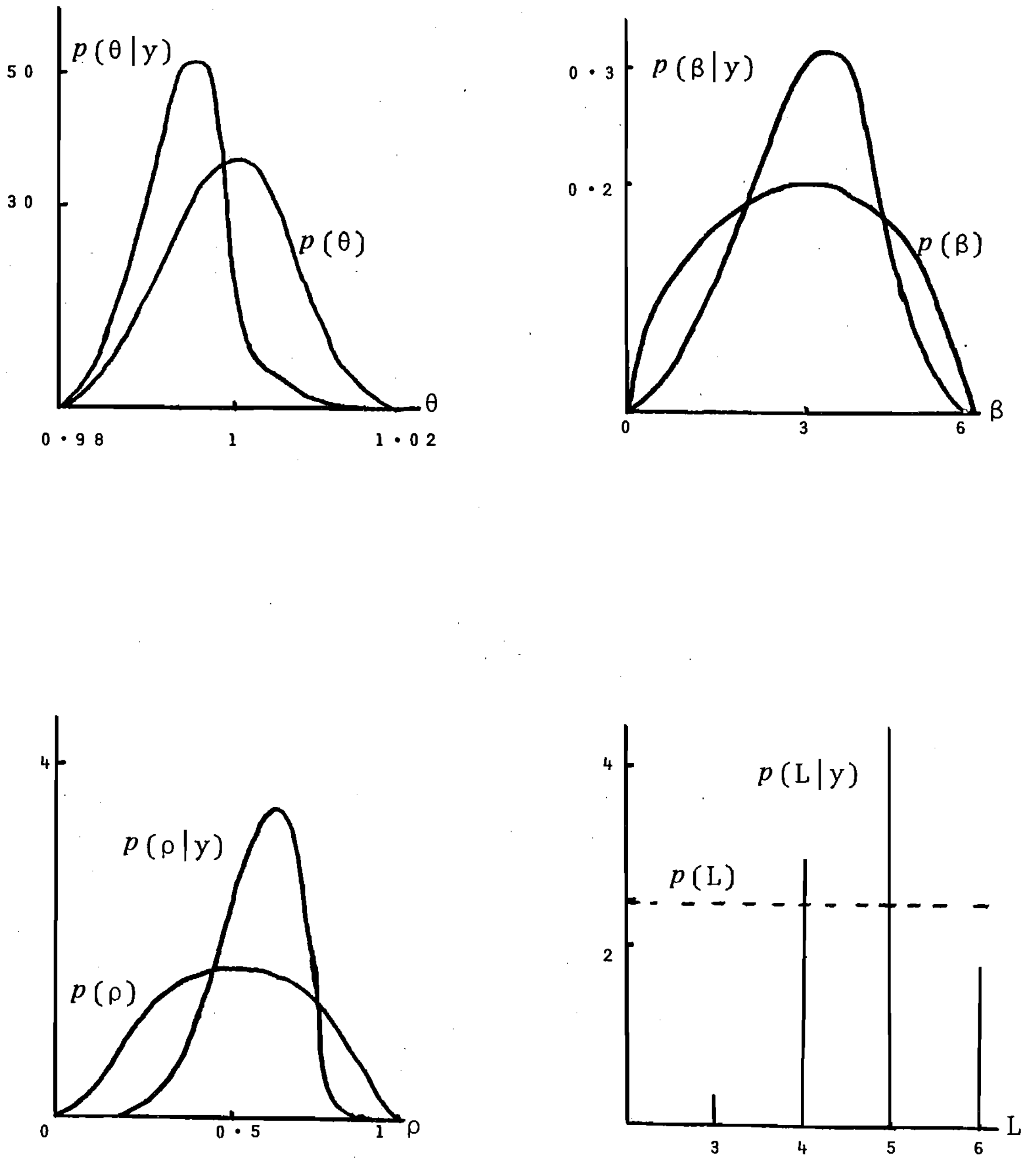


Figure VIII.VIII.16: Part 4, S_1 (Tight prior)

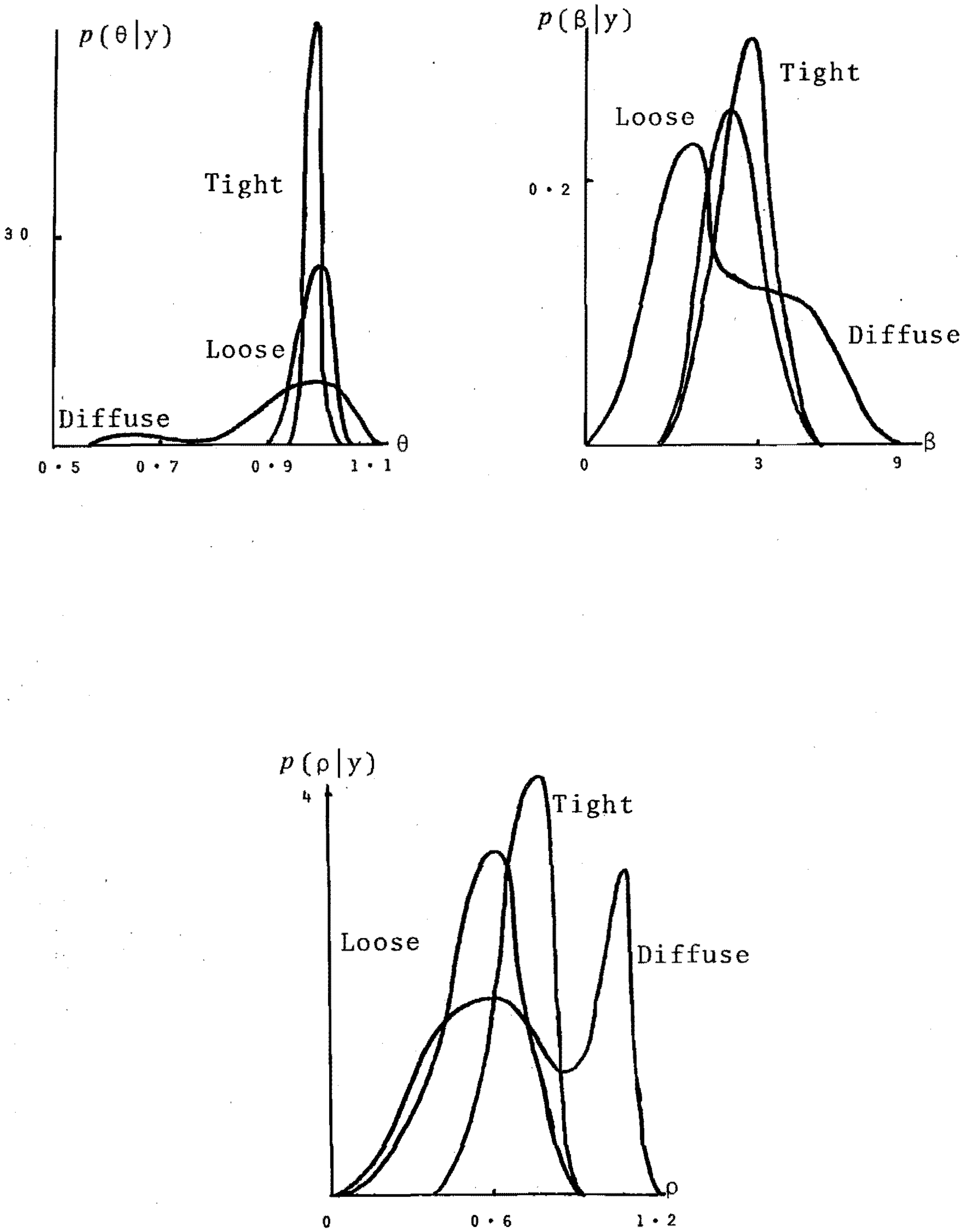


Figure VIII.VIII.17: Part 4, S_1 (Posteriors) p.m.f.'s are compared with their corresponding priors, and then with each other. For Part (1) of the study, the joint posterior density contours for $p(\theta,\beta|y)$ are shown for three different prior p.d.f.'s, by way of illustration.

In fact, plots of all the marginal posterior p.d.f.'s on which the tabulated results are based were generated and are available. However, those shown are indicative of the others, and represent the "most probable" models in each part of the study.

Presentation of full posterior densities is an important feature of Bayesian analysis, as summary statistics (such as moments) can often be misleading. This is expecially apparent in Figures VIII.VIII.9 and VIII.VIII.14 where bi-modal densities are depicted. In this case, one might question the use[18] of a quadratic loss function (which implies the mean of the posterior p.d.f. as a point estimate of the parameter), especially for the parameter ρ.

(3) <u>Appraisal of the Bayes-Almon Estimator</u>

The main objective of this applied study is to evaluate the Bayesian estimation procedure outlined in Chapter VII. Of special interest are those features of this procedure that are <u>not</u> incorporated in the estimator proposed by Zellner and Williams. These features include the treatment of L as a discrete random variable, allowance for serially correlated disturbances, and the use of posterior probabilities for various types of model-selection.

---

18. Bi-modal posterior p.d.f.'s are reported in other applied Bayesian studies. For example, see Zellner and Geisel (1970), and Geisel, <u>op.cit</u>., p.100.

The results obtained clearly illustrate the type of information that may be gained by using this Bayesian procedure.

However, there are several features of the results which are common to some other applied Bayesian studies, and these may cause concern in some cases.

The computational cost of this type of Bayesian analysis is very high, especially if numerical integration must be used to analyse the posterior p.d.f.'s, and in this case the size (in terms of number of parameters) of the models that may be studied is very limited.

Further, the introduction of prior information explicitly into the estimation may be a mixed blessing. If this information is such that the mean of the prior p.d.f. and the true value of the parameter are quite close, and if the data do not dominate this information, then the point estimates so obtained may have lower M.S.E. than the corresponding O.L.S. estimates.[19] However, the estimates could be very badly biased and/or have low M.S.E. if the prior mean is substantially different from the true parameter value, and if the prior variance is sufficiently small. Of course, no Bayesian denies that care should be taken with the specification of the prior p.d.f.'s.

Thirdly, there may be considerable variations in the posterior probabilities of the different models as the sample period is extended, or as different prior p.d.f.'s are tested.[20] Thus if terminal actions are based on a weighted average, $\bar{M}$, of all of the models in question,

---

19. Recall the discussion of Chapter III.
20. This is not a major problem in the present study.

these weights being the posterior probabilities, then the dominant models in $\bar{M}$ may differ from period to period.

Fourthly, there is always a danger that the "true" model (which generated the sample) is not among those being tested. This is a very real possibility in the present study. If this is so, then the posterior probabilities for the different models are incorrect, though the posterior odds taken pairwise across the model space are still valid.

Fifthly, a model with relatively low posterior probability may predict more accurately in a particular period than does a model with higher posterior probability. This is frequently the case in this study.[21] To some extent this may again suggest that the true model is not among those tested. However, it is to be expected that in any particular instance some incorrect model will predict more successfully than will the true model. The more models are being considered, the more likely is this situation. Further, only point predictions are computed in the present study. It may be that when a model with low posterior probability provides apparently accurate point predictions, in fact these predictions have large variances. Thus, it may be unfair to judge on this issue in the absence of interval forecasts.[22] Again, a strict Bayesian would use $\bar{M}$ when predicting.

Of course, all of the above comments may be relevant to any applied Bayesian study.

---

21. Similar results are reported by Geisel, op.cit., p.58 ff.
22. Ideally, of course, a Bayesian would wish to present the full predictive p.d.f. for the period in question.

However, the results obtained in this Chapter are encouraging in many respects. Despite the heavy computational cost (unless natural-conjugate or "diffuse" prior p.d.f.'s are used), the analysis can be formalized to a degree that is impossible if a classical approach is used.

For example, allowing L to be a discrete random variable means that estimates of other parameters incorporate the uncertainty surrounding L. Further, posterior probabilities facilitate comparisons of models based on different values of L, different polynomial degrees, and different sets of restrictions on the weights in the distributed lag(s). Formal model comparisons of this type generally are not possible in the classical Almon estimation procedure.

Finally, although allowance for serially correlated disturbances can be made in the classical Almon estimator, in the Bayesian approach used here such an allowance is more flexible in that prior information about serial correlation is incorporated in the same way as is such information about other parameters.

As is emphasised in Chapter VII, the generalized Bayes-Almon estimator used here relies critically on the imposition of P independent linear restrictions on the w_k's in VIII.II.1 when a polynomial of degree P is used. The risks involved when imposing such restrictions were also noted in Chapter VII. However, such restrictions are encountered frequently in applied studies involving the Almon estimator, so in this respect the present study is not alone. Further, in the Bayesian version, at least the effects of imposing different restrictions on the w_k's may be analysed formally by means of B.P.O. analysis.

In summary, although other Bayesian and non-Bayesian methods of estimating finite distributed lag models are available, if the Almon method is used then there are several important advantages in taking a Bayesian approach. Although there are computational limitations, many of the unfortunate difficulties arising with the classical Almon estimator are overcome.

(4) Economic Implications

The basic economic relationship used in the study is extremely simple, for reasons already noted, and it is acknowledged that in fact a more complex relationship may be more appropriate. In particular, we have abstracted from the problems of simultaneity bias by considering only a single equation.

Accordingly, only limited economic implications may be drawn from the study. However, the results strongly favour a lag shape exhibiting rapid initial decay and slow final decay over a period of four or five quarters. Further, the results suggest that the lag relationship is relatively free of "leakages", since there is strong evidence that the sum of the lag weights is close to unity. Generally, these results agree with the findings of Deane et al., on which this study is based.

Finally, the influence of the availability of trading bank credit (as tested via a dummy variable) on payments for c.i.f. imports is open to some doubt, in view of the rather large posterior variances generally exhibited by the associated parameter. Of course, this last conclusion may be changed if more complex relationships were analysed.

IX. CONCLUDING REMARK

The main purpose of this Chapter is to evaluate the general Bayesian verison of the Almon estimator for finite distributed lag models, as introduced in Chapter VII.

Although the study is necessarily limited in scope, it seems clear that this generalized M.E.L. estimator is capable of overcoming most of the methodological difficulties associated with its classical counterpart, though generally at greater computational cost.

CHAPTER IX

SUMMARY AND CONCLUSIONS

The purpose of this thesis is to investigate several related aspects of Bayesian inference in econometrics and to illustrate their application. To some extent the study was motivated by a degree of dissatisfaction with the inability of classical methods of inference to handle a number of interesting econometric problems. In particular, problems relating to model comparisons and to the analysis of finite distributed lag models raise methodological issues which cannot be resolved adequately by the existing classical inferential tools. In contrast to this, the available Bayesian techniques are admirably suited to handling some of these problems, as is demonstrated in this study.

The sampling properties of Bayes (or M.E.L.) estimators have received relatively little attention in the econometric literature, and Chapter III is devoted to an investigation of such properties for the natural-conjugate Bayes estimator in the multiple regression model. These properties are compared with those of the O.L.S. estimator, using matrix M.S.E. as a basis for comparison. The results obtained provide some interesting insights into the sampling properties of this M.E.L. estimator, and the analysis could be extended by taking account of pre-testing bias and /or using the weaker M.S.E. criterion proposed by Wallace (1972).

The results obtained in Chapter III form part of the motivation for the analysis in Chapter IV, where a M.E.L. procedure is proposed for seasonally adjusting economic time-series. This analysis provides a natural extension of a well-known classical method based on dummy variables and suggested by Jorgenson (1964). The proposed analysis uses conventional M.E.L. estimation and B.P.O. analysis. The contents of this Chapter exhibit yet another application of these tools to a common inference problem in econometrics, and this material forms a partial link between Chapter III and the latter part of the thesis.

The rest of the study is devoted to various aspects of distributed lag models, a brief survey of recent Bayesian contributions in this area being provided in Chapter V. The analysis of such models raises substantial difficulties for the classical econometrician. Here, the flexibility with which a Bayesian approach allows uncertain <u>a priori</u> information to be used when estimating parameters is especially helpful. Further, distributed lag models give rise to well-defined model comparison problems for which the classicist has no generally applicable tool. In such cases, B.P.O. analysis provides the ideal means of discriminating among different specifications, and a specific problem is discussed and analysed in Chapter VI.

Several methodological difficulties are associated with the classical Almon estimator for finite distributed lag models in practice, these centering on the use of exact <u>a priori</u> restrictions, and model comparisons. These and related difficulties are analysed in Chapter VII, and a recent important Bayesian contribution by Zellner and Williams (1973) is discussed. The rest of that Chapter is devoted to a theoretical discussion of various extensions of their analysis, special account being taken of prior information; the maximum lag length; serial correlation; and model comparisons. The result is a Bayesian estimator which is as flexible as its classical counterpart, yet is able to handle the above-mentioned problems far more easily and formally . In many ways this Chapter demonstrates some of the most useful aspects of Bayesian inference in econometrics.

In Chapter VIII the theory of Chapter VII is applied to some New Zealand data, and the results obtained illustrate the advantages and scope of the Bayesian analysis. However, this study also illustrates the high computational costs associated with such analyses if proper prior p.d.f.'s other than the natural-conjugate densities are used. At present, if the prior information cannot be summarized adequately by a diffuse or natural-conjugate prior p.d.f., then a Bayesian analysis is prohibitively expensive in parameter spaces of dimension in excess of about four.

However, despite this severe practical limitation, one major conclusion that may be drawn from the thesis is that there are a number of important and interesting econometric problems where a Bayesian analysis is helpful. Although only a few such problems are dealt with in the thesis, the results obtained demonstrate some of the features of Bayesian inference as they relate to estimation, prediction, prior information, nuisance parameters, and specification analysis. In doing so they serve to reinforce the results of earlier related work.

Clearly, much remains to be done, and it is to be hoped that in the future Bayesian methodology will be explored and applied more widely in the context of econometric problems. In recent years the necessary foundations have been laid, and although the Bayesian approach to econometric inference cannot provide all of the answers, already it has proved most helpful in several situations in which classical analysis is inadequate.

The burden of writing this thesis has been eased considerably by moral and financial assistance from a variety of sources, and for this I am extremely grateful.

First and foremost I owe a great deal to my supervisor, Professor Tony Rayner, who originally suggested that I work in the general area of Bayesian inference, and who subsequently provided constant and enthusiastic encouragement, as well as much needed constructive criticism of my work.

It is a pleasure to acknowledge the financial assistance provided by a University Grants Committee Postgraduate Scholarship and a William Georgetti Scholarship. In addition, Professor A. D. Brownlie kindly provided general and financial help by arranging for teaching assignments, excellent office accommodation, and free access to computer and typing facilities. I am most grateful for this support.

The problems associated with geographical isolation from other researchers in my area of interest were overcome to some extent, but only by the friendly co-operation of a number of correspondents, who promptly and generously supplied many unpublished papers, without which my task would have been impossible. In this respect I would like to thank Dr W. E. Griffiths, and Professors K. M. Gaver, M. S. Geisel, E. E. Leamer, T. A. Rothenberg, R. J. Shiller and C. A. Sims.

I am especially grateful to Professor A. Zellner, who not only provided a wealth of unpublished material, but who also offered valuable criticism of earlier drafts of Chapter VI of the thesis.

Past and present members of the staff of the Department of Economics at the University of Canterbury helped by suffering through lengthy discussions on matters Bayesian, and in particular I would like to thank Richard Manning, John George and Peter Morgan.

The data used in Chapter VIII of the thesis were made available by Rod Deane and John Gallacher of the Reserve Bank of New Zealand. Their prompt and willing attention was greatly appreciated.

The typing of most of the final version of the thesis and much of the earlier drafting was handled most cheerfully and efficiently by Lynne Worrill, whose enthusiasm for this onerous task stands to her credit. I am also very grateful to Nicky Hunter and Margaret Sewell for their help with the typing of the final manuscript.

Finally, the greatest debt of all is to my wife, Lorraine, who unselfishly withstood the demands and crises associated with the thesis, and still found the energy to provide essential encouragement.

REFERENCES

Allen, D.M. (1971), "Mean Square Error of Prediction as a Criterion for Selecting Variables", Techonometrics, 13 pp. 469-475.

Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures", Econometrica, 33, pp. 178-196.

Almon, S. (1968),"Lags Between Investment Decisions and Their Causes", Review of Economics and Statistics, 50, pp. 193-206.

Bancroft, T.A. (1944), "On Biases in Estimation Due to the Use of Preliminary Tests of Significance", Annals of Mathematical Statistics, 15, pp. 190-204.

Bartlett, M.S. (1965), "R.A. Fisher and the Last Fifty Years of Statistical Methodology", Journal of the American Statistical Association, 60, pp. 395-409.

Bock, M.E., G.G. Judge, and T.A. Yancey (1973), "Some Comments on Estimation in Regression After Preliminary Tests of Significance", Journal of Econometrics, 1, pp. 191-200.

Box, G.E.P., and G.M. Jenkins (1971), Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day.

Box, G.E.P. and G.C. Tiao (1973), Bayesian Inference in Statistical Analysis. Reading Massachusetts, Addison-Wesley.

Chetty, V.K. (1971), "Estimation of Solow's Distributed Lag Models", Econometrica, 39, pp. 99-117.

Christ, C.F. (1966), Econometric Models and Methods. New York, Wiley.

Cleveland, W.S. (1971), "Estimation of Parameters in Distributed Lag Econometric Models". Department of Statistics, University of North Carolina at Chapel Hill, Mimeo. Series No. 792.

Cochrane, D. and G.H. Orcutt (1949), "Applications of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", Journal of the American Statistical Association, 44, pp. 32-61.

Cohen, M., R. Gillingham and D. Heien (1973), "A Monte Carlo Study of Complex Finite Distributed Lag Structures", Annals of Economic and Social Measurement, 2, pp. 53-63.

Cooper, J.P. (1972), "Two Approaches to Polynomial Distributed Lag Estimation: An expository Note and Comment", American Statistician, 26, pp. 32-35.

Courville, L.C. and M.S. Geisel (1971), "Comparison of Alternative Lag Distributions and Error Structures". Graduate School of Industrial Administration, Carnegie - Mellon University, Unpublished Manuscript.

Deane, R.S. and M.A. Lumsden (1972), "An Econometric Approach to Forecasting New Zealand's Imports". Reserve Bank of New Zealand, Research Paper No. 5.

Deane, R.S., M.A. Lumsden, and A.B. Sturm (1972), "Some Simulation Experiments with a New Zealand Model". In R.S. Deane (ed.), "A New Zealand Model: Structure, Policy Uses and Some Simulation Results". Reserve Bank of New Zealand, Research Paper No. 8.

Dhrymes, P.J. (1969), "Efficient Estimation of Distributed Lags With Autocorrelated Errors", International Economic Review, 10, pp. 41-67.

Dhrymes, P.J. (1971), "On the Strong Consistency of Estimators for Certain Distributed Lag Models with Autocorrelated Disturbances", International Economic Review, 12, pp. 329-343.

Dhrymes, P.J. (1971a), Distributed Lags. San Francisco, Holden-Day.

Dhrymes, P.J., L.R. Klein, and K. Steiglitz (1970), "Estimation of Distributed Lags", International Economic Review, 11, pp. 235-250.

Fair, R.C. and D.M. Jaffee (1971), "A Note On the Estimation of Polynomial Distributed Lags". Econometric Research Program, Princeton University, Research Memorandum No. 120.

Feyerabend, P. (1965), "Problems of Empiricism". In R. Colodny (ed.), Beyond the Edge of Certainty. Edgewood Cliffs New Jersey, Prentice-Hall.

Finetti, B. de (1937), "La Prévision: Ses Lois Logiques, Ses Sources Subjectives", Annales de l'Institut Henri Poincaré, 7, pp. 1-68.

Finetti, B. de (1972), Probability, Induction and Statistics. New York, Wiley.

Fisher, I. (1937), "A Note On a Short-Cut Method for Calculating Distributed Lags", Bulletin de l'Institut International de Statistique, 29, pp. 323-327.

Fleiss, J.L. (1966), "A Note On the Expectation of the Reciprocal of a Random Variable", American Statistician, 20, p. 25.

Gaver, K.M. (1974), "Comparing Multiple Equation Models". Graduate School of Industrial Administration, Carnegie-Mellon University, Unpublished Ph.D. Dissertation.

Gaver, K.M. and M.S. Geisel (1973), "Discriminating Among Linear Models with Interdependent Disturbances". Graduate School of Management, University of Rochester, Unpublished Manuscript.

Gaver, K.M. and M.S. Geisel (1974), "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods". In P. Zarembka (ed.), Frontiers of Econometrics. New York, Academic Press.

Geisel, M.S. (1970), "Comparing and Choosing Among Parametric Statistical Models: A Bayesian Analysis with Macro-economic Applications". Graduate School of Business, University of Chicago, Unpublished Ph.D. Dissertation.

Giles, D.E.A. (1973), "Recent Contributions to the Estimation of Distributed Lags". Department of Economics, University of Canterbury, Discussion Paper No. 7309.

Giles, D.E.A. (1974), "The Almon Estimator and Serially Correlated Disturbances: A Comment and Application to a New Zealand Imports Payments Function", New Zealand Economic Papers, 8, pp. 138-150.

Giles D.E.A. (1974a), "The Restricted P.A.E.: Tests of Significance and the Average Lag Length". Department of Economics, University of Canterbury, Unpublished Manuscript.

Goldberger, A.S. (1964), Econometric Theory. New York, Wiley.

Graybill, F.A. (1961), An Introduction to Linear Statistical Models, Vol. 1, New York, McGraw-Hill.

Griffiths, W.E. (1973), "A New Look at Ridge Regression: Biased Estimators with Lower Mean Square Error". Paper Presented at the Third Conference of Economists, Adelaide South Australia.

Griliches, Z. (1967), "Distributed Lags: A Survey", Econometrica, 35, pp. 16-39.

Groot, M.H. de (1970), Optimal Statistical Decisions. New York, Wiley.

Hadley, G. (1964), Nonlinear and Dynamic Programming. Reading Massachusetts, Addison-Wesley.

Haitovsky, Y. (1969), "A Note On the Maximization of $\bar{R}^2$", American Statistician, 23, pp. 20-21.

Hall, R.E. and R.C. Sutch (1968), "A Flexible Infinite Distributed Lag". Institute of Business and Economic Research, University of California at Berkeley, Working Paper No. 124.

Hamming, R.W. (1962), Numerical Methods for Scientists and Engineers. New York, McGraw-Hill.

Hildreth, C. and J.Y. Lu (1960), "Demand Relations With Autocorrelated Disturbances", Agricultural Experiment Station, Michigan University, Technical Bulletin No. 276.

Jeffreys, H. (1961), Theory of Probability, 3rd Revised Edition. Oxford, Clarendon.

Jenkins, D.M. and D.G. Watts (1968), Spectral Analysis and its Applications. San Francisco, Holden-Day.

Johnston, J. (1972), Econometric Methods, 2nd Edition. New York, McGraw-Hill.

Jorgenson, D.W. (1964), "Minimum Variance, Linear, Unbiased Seasonal Adjustment of Economic Time-Series", Journal of the American Statistical Association, 59, pp.681-724.

Jorgenson, D.W. (1966), "Rational Distributed Lag Functions", Econometrica, 34, pp. 135-149.

Jorgenson, D.W. (1967), "Seasonal Adjustment of Data for Econometric Analysis", Journal of the American Statistical Association, 62, pp. 137-140.

Judge, G.G., T.A. Yancey, and M.E. Bock (1973), "Properties of Estimators After Preliminary Tests of Significance When Stochastic Restrictions are Used in Regression", Journal of Econometrics, 1, pp. 29-47.

Judge, G.G., M.E. Bock and T.A. Yancey (1974), "Post Data Model Evaluation", Review of Economics and Statistics, 56, pp. 245-253.

Kadiyala, K.R. (1968), "A Transformation Used to Circumvent the Problem of Autocorrelation", Econometrica, 36, pp. 93-96.

Koyck, L.M. (1954), Distributed Lags and Investment Analysis. Amsterdam, North-Holland.

Kuhn, T. (1962), The Structure of Scientific Revolution. Chicago, University of Chicago Press.

Kyburg, H.E. and H.E. Smokler (1964), Studies in Subjective Probability. New York, Wiley.

Leamer, E.E. (1970), "Another Class of Prior for Distributed Lag Analysis". Harvard Institute of Economic Research, Discussion Paper No. 147.

Leamer, E.E. (1972), "Distributed Lag Analysis With Informative Priors", Econometrica, 40, pp. 1059-1082.

Leamer, E.E. (1973), "Multicollinearity: A Bayesian Interpretation", Review of Economics and Statistics, 55, pp. 371-380.

Lempers, F.B. (1971), Posterior Probabilities of Alternative Linear Models. Rotterdam, Rotterdam University Press.

Lindley, D.V. (1961), "The Use of Prior Probability Distributions in Statistical Inferences and Decisions". In J. Neyman (ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley, University of California Press.

Lindley, D.V. (1965), Introduction to Probability and Statistics From a Bayesian Viewpoint. London, Cambridge University Press.

Lovell, M.C. (1963), "Seasonal Adjustment of Economic Time Series", Journal of the American Statistician Association, 58, pp. 993-1010.

Lovell, M.C. (1966), "Alternative Axiomatizations of Seasonal Adjustment", Journal of the American Statistical Association, 61, pp. 800-802.

Maddala, G.S. (1971), "Some Notes On Discrimination Between Different Distributed Lag Models". Centre for Operations Research and Econometrics, Université Catholique de Louvain, Discussion Paper No. 7109.

Maddala, G.S. and A.S. Rao (1971), "Maximum Likelihood Estimation of Solow's and Jorgenson's Distributed Lag Models", Review of Economics and Statistics,53,pp.80-88.

McCallum, B.T. (1970), "Artificial Orthogonalization in Regression Analysis", Review of Economics and Statistics, 52, pp. 110-113.

McNown, R.F. (1971), "A Simulation Study of the Small-Sample Properties of Several Distributed Lag Estimators". Department of Economics, University of California at San Diego, Unpublished Ph.D. Dissertation.

Mills, F.C. (1955), Statistical Methods. New York, Holt, Rinehart and Winston.

Nerlove, M. (1972), "Lags in Economic Behaviour", Econometrica, 40, pp. 221-252.

Neumann, J. von and O. Morgenstern (1947), Theory of Games and Economic Behaviour. Princeton, Princeton University Press.

Palm, F. (1972), "On Mixed Prior Distributions and Their Application in Distributed Lag Models". Centre for Operations Research and Econometrics, Université Catholique de Louvain, Discussion Paper No. 7222.

Pease, M.C. (1965), Methods of Matrix Algebra. New York, Academic Press.

Plackett, R.L. (1966), "Current Trends in Statistical Inference", Journal of the Royal Statistical Society, A, 129, pp. 249-267.

Popper, K.R. (1965), The Logic of Scientific Discovery. New York, Science Editions.

Raiffa, H. and R. Schlaifer (1961), Applied Statistical Decision Theory. Boston, Harvard University.

Ramsey, F.P. (1950), "Truth and Probability". In R.B. Braithwaite (ed.), The Foundations of Mathematics and Other Logical Essays. New York, Humanities Press.

Rand Corporation (1950), A Million Random Digits and One Hundred Thousand Deviates. Santa Monica, Rand Corporation.

Rao, A.S. (1971), "Likelihood Methods for the Estimation of Distributed Lag Models". Graduate School of Management, University of Rochester, Unpublished Ph.D. Dissertation.

Rao, C.R. (1965), Linear Statistical Inference and its Applications. New York, Wiley.

Renyi, A. (1970), Foundations of Probability. San Francisco, Holden-Day.

Rothenberg, T.J. (1969), "The Bayesian Approach and Alternatives in Econometrics - II". Paper Presented at the Winter Meetings of the Econometric Society, New York.

Savage, L.J. (1954), Foundations of Statistics. New York, Wiley.

Savage, L.J. (1961), "The Foundations of Statistics Reconsidered". In J. Neyman (ed.), Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability, Vol. 1. Berkeley, University of California Press.

Schmidt, P. (1974), "A Note On Theil's Minimum Standard Error Criterion When the Disturbances are Autocorrelated", Review of Economics and Statistics, 56, pp. 122-123.

Schmidt, P. and R.N. Waud (1973), "The Almon Lag Technique and the Monetary versus Fiscal Policy Debate", Journal of the American Statistical Association, 68, pp. 11-19.

Sclove, S.L. (1968), "Improved Estimators for Coefficients in Linear Regression", Journal of the American Statistical Association, 63, pp. 596-606.

Scott, D.B. and S.R. Timms (1966), Mathematical Analysis. London, Cambridge University Press.

Searle, S.R. (1971), Linear Models. New York, Wiley.

Schiller, R.J. (1973), "A Distributed Lag Estimator Derived from Smoothness Priors", Econometrica, 41, pp. 775-788.

Silvey, S.D. (1970), Statistical Inference. Harmondsworth, Penguin.

Sims, C.A. (1971), "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite Dimensional", Annals of Mathematical Statistics, 42, pp. 1622-1636.

Sims, C.A. (1971a), "Discrete Approximations to Continuous Time Distributed Lags in Econometrics", Econometrica, 39, pp. 545-564.

Sims, C.A. (1972), "The Role of Approximate Prior Restrictions in Distributed Lag Estimation", Journal of the American Statistical Association, 67, pp. 169-175.

Sims, C.A. (1972a), "Seasonality in Regression". Centre for Economic Research, Department of Economics, University of Minnesota, Discussion Paper No. 23.

Sims, C.A. (1973), "Distributed Lags". Centre for Economic Research, Department of Economics, University of Minnesota, Discussion Paper No. 28.

Smith, A.F.M. (1973), "A General Bayesian Linear Model", Journal of the Royal Statistical Society, B, 53, pp. 67-75.

Solow, R.M. (1960), "On a Family of Distributed Lags", Econometrica, 28, pp. 393-406.

Steiglitz, K. and L.E. McBride (1965), "A Technique for the Identification of Linear Systems", I.E.E.E. Transactions on Automatic Control, AC-10, pp. 461-464.

Stein, C. (1960), "Multiple Regression". In I. Olkin (ed.), Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling. Stanford, Stanford University Press.

Theil, H. (1961), Economic Forecasts and Policy. Amsterdam, North-Holland.

Theil, H. (1971), Principles of Econometrics. Amsterdam, North-Holland.

Thornber, E.H. (1966), "Applications of Decision Theory to Econometrics". Department of Economics, University of Chicago, Unpublished Ph.D. Dissertation.

Thornber, E.H. (1967), "Finite Sample Monte Carlo Studies: An Autoregressive Illustration", Journal of the American Statistical Association, 62, pp. 801-818.

Tiao, G.C. and G.E.P. Box (1973), "Some Comments on 'Bayes' Estimators", American Statistician, 27, pp. 12-14.

Tiao, G.C. and A. Zellner (1964), "Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis", Biometrika, 51, pp. 219-230.

Toro-Vizcarrondo, C. and T.D. Wallace (1968), "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression", Journal of the American Statistical Association, 63, pp. 558-572.

Trivedi, P.K. (1970), "A Note On the Application of Almon's Method of Calculating Distributed Lag Coefficients", Metroeconomica, 22, pp. 281-286.

Tsurumi, H. (1973), "Bayesian Estimation and Comparison of Gamma and Poisson Distributed Lag Models". Department of Economics, Queens University, Discussion Paper No.127.

Wald, A. (1950), Statistical Decision Functions. New York, Wiley.

Wallace, T.D. (1972), "Weaker Criteria and Tests for Linear Restrictions in Regression", Econometrica, 40, pp. 689-698.

Wallace, T.D. and C. Toro-Vizcarrondo (1969), "Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression", Journal of the American Statistical Association, 64, pp. 1649-1663.

Wallis, K.F. (1969), "Some Recent Developments in Econometrics: Dynamic Models and Simultaneous Equation Systems", Journal of Economic Literature, 7,pp. 771-796.

Zellner, A. (1969), "The Bayesian Approach and Alternatives in Econometrics - I". Paper presented at the Winter Meetings of the Econometric Society, New York.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics. New York, Wiley.

Zellner, A. (1972), "On Assessing Informative Prior Distributions for Regression Coefficients". H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago, Unpublished Manuscript.

Zellner, A. and M.S. Geisel (1970), "Analysis of Distributed Lag Models With Applications to Consumption Function Estimation", Econometrica, 38, pp. 865-888.

Zellner, A. and F. Palm (1974), "Time Series Analysis and Simultaneous Equation Models", Journal of Econometrics, 2, pp. 17-54.

Zellner, A. and G.C. Tiao (1964), "Bayesian Analysis of the Regression Model with Autocorrelated Errors", Journal of the American Statistical Association, 59, pp. 763-768.

Zellner, A. and W. Vandaele (1972), "Bayes-Stein Estimators for κ-Means, Regression and Simultaneous Equation Models". H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago, Unpublished Manuscript.

Zellner, A. and A.D. Williams (1973), "Bayesian Analysis of the Federal Reserve - M.I.T. - Penn Model's Almon Lag Consumption Function", Journal of Econometrics, 1, pp. 267-299.

APPENDIX I

SOME USEFUL RESULTS

Theorem A.1: If P is a non-singular matrix, then P'AP is p.(s).d. according to whether A is p.(s).d..

Proof: Searle (1971), p.36.

Theorem A.2: Under the assumptions of Theorem A.1, the matrix A^{-1} is p.d.s. iff A is p.d.s..

Proof: Apply Theorem A.1 with $P=A^{-1}$, and with P=A.

Theorem A.3: Let G and H be p.(s).d.s. matrices of the same order. Then (G+H) is at least p.s.d.s., and is p.d.s. if at least one of G or H is p.d.s..

Proof: The symmetry result is trivial. Further, $\eta' H\eta \geq 0$, $\eta' G\eta \geq 0$, for all $\eta \neq 0$, with strict inequality iff the matrix concerned is p.d..

Thus, $\eta'(G+H)\eta = \eta' G\eta + \eta' H\eta \geq 0$, with strict inequality iff at least one of G and H is p.d..

Theorem A.4: If A is p.d. then A is non-singular.

Proof: Searle, *op.cit*. p.36.

Theorem A.5: If A is p.s.d. then A is singular.

Proof: Searle, op.cit., p.36.

Theorem A.6: Let A, B and C be matrices of such dimensions that CAB is defined, and such that B and C are non-singular. Then rank(CAB) = rank(A).

Proof: Johnston (1972), p.100.

Theorem A.7: $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$

Proof: Graybill (1961), p.2.

Theorem A.8: Let U be an (nxn) matrix function of scalar z. Then, if m is a positive integer,

$$(\partial U^{-m}/\partial z) = -\{U^{-m}U'U^{-1}+U^{-(m-1)}U'U^{2}+ --- + --- + U^{-1}U'U^{-m}\},$$

where a prime here denotes a first derivative.

Proof: Pease (1965), pp.166-168.

Theorem A.9: If η is an arbitrary non-zero k-vector, then $(\eta\eta')$ is a p.s.d.s. matrix.

Proof: $$(\eta\eta') = \begin{pmatrix} \eta_1^2 & \eta_1\eta_2 & \cdots & \eta_1\eta_k \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \eta_k\eta_1 & \cdots & \cdots & \eta_k^2 \end{pmatrix}$$

which is symmetric. Further, if x is any non-zero k-vector, then

$$x'(\eta\eta')x = (\eta'x)'(\eta'x), \text{ where } \eta'x \text{ is scalar.}$$

$$\text{So, } x'(\eta\eta')x = [\sum_{1}^{k}\eta_i x_i]^2 \geq 0,$$

so $(\eta\eta')$ is p.s.d.s..

Theorem A.10: Let the vector x be distributed $N(\mu,V)$. Then $x'Dx$ is non-central χ^2 with rank(D) degrees of freedom and with non-centrality parameter $\rho=\frac{1}{2}\mu'D\mu$, iff DV is idempotent.

Proof: Searle, *op.cit.*, pp.57-58.

Theorem A.11: Let x_1 and x_2 be normally distributed random variables. Then x_1 and x_2 are independent iff $\mathrm{cov.}(x_1,x_2)=0$.

Proof: Searle, *op.cit.*, p.47.

Theorem A.12: Any symmetric matrix A of order n and rank r can be written as LL', where L is $(n\times r)$ of rank r.

Proof: Searle, *op.cit.*, p.37.

Theorem A.13: If f is a continuous function of x, and $\operatorname*{plim}_{n\to\infty} (x_n) = x$, then $\operatorname*{plim}_{n\to\infty} [f(x_n)] = f(x)$.

Proof: Rao (1965), p.104.

Theorem A.14: (Jensen's Inequality):

If x is a random variable such that $\mathcal{E}(x)=\mu$, and if $f(x)$ is a convex function, then $\mathcal{E}[f(x)] \geq f[\mathcal{E}(x)]$, with equality iff x has a degenerate distribution at μ.

Proof: Rao, op.cit., pp. 46-47.

Theorem A.15: $\sum_{r=0}^{n} r = n(n+1)/2$; $\sum_{r=0}^{n} r^2 = n(n+1)(2n+1)/6$;

$$\sum_{r=0}^{n} r^3 = n^2(n+1)^2/4,$$

for positive integer n.

Proof: Scott and Tims (1966), pp. 40-41.

<u>Theorem A.16</u>: $\sum_{r=0}^{n} r^4 = n(n+1)(2n+1)(3n^2+3n-1)/30$, for positive integer n.

<u>Proof</u>: $(r+1)^5 - r^5 = 5r^4 + 10r^3 + 10r^2 + 5r + 1$

So,

$$(n+1)^5 - n^5 = 5n^4 + 10n^3 + 10n^2 + 5n + 1$$
$$n^5 - (n-1)^5 = 5(n-1)^4 + 10(n-1)^3 + 10(n-1)^2 + 5(n-1) + 1$$
$$\vdots$$
$$2^5 - 1^5 = 5^4 + 10^4 + 10^2 + 5 + 1.$$

Hence,

$$(n+1)^5 - 1^5 = 5\sum_{r=1}^{n} r^4 + 10\sum_{r=1}^{n} r^3 + 10\sum_{r=1}^{n} r^2 + 5\sum_{r=1}^{n} r + n$$

and,

$$\sum_{r=0}^{n} r^4 = \sum_{r=1}^{n} r^4 = \{(n+1)^5 - 1 - 10\sum_{r=1}^{n} r^3 - 10\sum_{r=1}^{n} r^2 - 5\sum_{r=1}^{n} r - n\}/5$$

$$= \{(n+1)^5 - (n+1) - 10[n^2(n+1)/4] - 10[n(n+1)(2n+1)/6] - 5[n(n+1)/2]\}/5$$

$$= (n+1)\{6(n+1)^4 - 6 - 15n^2(n+1) - 10n(2n+1) - 15n\}/30$$

$$= (n+1)(6n^4 + 9n^3 + n^2 - n)/30$$

$$= n(n+1)(2n+1)(3n^2 + 3n + 1)/30$$

<u>Theorem A.17</u>: If x and y are positive independent random variables with finite means, then

$$\mathcal{E}(x/y) > \mathcal{E}(x)/\mathcal{E}(y)$$

<u>Proof</u>: Fleiss (1966), p. 25.

<u>Theorem A.18</u>: Let x be any random vector with mean μ and covariance matrix V. Then

$$\mathcal{E}(x'Ax) = \text{tr.}(AV) + \mu'A\mu \ .$$

<u>Proof</u>: Searle, <u>op.cit</u>., p. 55.

APPENDIX II

SOME RESULTS FOR THE ONE-REGRESSOR MODEL

Consider some of the analysis in Part (2) of Section IV of Chapter III, as it relates to the simple one-regressor model used in Section VI of that Chapter:

$$y_t = \beta x_t + \varepsilon_t \; ; \qquad t = 1, 2, \ldots, n$$

where β is a scalar parameter.

In this case, the matrix A^{-1} is replaced by the scalar, a, so in terms of the notation in Chapter III, ϕ is now replaced by a. Since (X'X) is now $\sum_t x_t^2$, also a scalar, (X'X) $\propto$ A is satisfied and so the Curve Décolletage is a line-segment on the β-axis.

Now,

$$B = \text{Bias}^2(\tilde{\beta}) = \beta^{*2}/a^2\,[(1/a) + \sum_t x_t^2]^2$$

so,

$$(\partial B/\partial a) = -2(\sum_t x_t^2)\beta^{*2}/(1 + a\sum_t x_t^2)^3$$

$$< 0,$$

and,

$$(\partial^2 B/\partial a^2) = 6\beta^{*2}(\sum_t x_t^2)/(1 + a\sum_t x_t^2)^4$$

$$> 0,$$

so that B is convex in a.

Also,

$$\lim_{a \to 0} (B) = \beta^{*2}; \quad \lim_{a \to 0} (\partial B/\partial a) = -2\beta^{*2}\sum_t x_t^2 < 0,$$

and

$$\lim_{a \to \infty} (B) = 0 \quad ; \quad \lim_{a \to \infty} (\partial B/\partial a) = 0.$$

Further,

$$V = \text{var.}\ (\tilde{\beta}) = \sigma^2(\sum_t x_t^2)/[(1/a) + \sum_t x_t^2]^2$$

so,

$$(\partial V/\partial a) = 2\sigma^2(\sum_t x_t^2)/[a^2(1/a + \sum_t x_t^2)^3]$$

$$> 0,$$

and,

$$(\partial^2 V/\partial a^2) = 2\sigma^2(\sum_t x_t^2)(1 - 2a\sum_t x_t^2)/(1 + a\sum_t x_t^2)^4$$

Thus, V has a point of inflexion at

$$a = (1/2\sum_t x_t^2),$$

and,

$$\lim_{a \to 0} (V) = 0 \quad ; \quad \lim_{a \to 0} (\partial V/\partial a) = 0$$

$$\lim_{a \to \infty} (V) = (\sigma^2/\sum_t x_t^2); \quad \lim_{a \to \infty} (\partial V/\partial a) = 0.$$

Finally, M = M.S.E. $(\tilde{\beta})$ has a minimum value at $a = (\beta^*/\sigma)^2$, and a point of inflexion at $a = (3/2)(\beta^*/\sigma)^2$; and,

$$\lim_{a \to 0} (M) = \beta^{*2} = \lim_{a \to 0} (B)$$

$$\lim_{a \to 0} (\partial M/\partial a) = -2\beta^{*2}\sum_t x_t^2 = \lim_{a \to 0} (\partial B/\partial a)$$

$$\lim_{a \to \infty} (M) = (\sigma^2/\sum_t x_t^2) = \lim_{a \to \infty} (V)$$

$$\lim_{a \to \infty} (\partial M/\partial a) = 0 = \lim_{a \to \infty} (\partial V/\partial a)$$

APPENDIX III

LAG DISTRIBUTION SHAPES

In this Appendix we present eight commonly encountered lag distribution shapes that are pertinent to Chapters VII and VIII. All but the first of these shapes can be described by some suitable polynomial of degree $P \geq N$, where N is the number of independent linear homogeneous restrictions imposed on the w_k's. (The first shape can be achieved only if P=N, _ceteris paribus_.)

The shapes are shown here in their most general form over the interval $[0,L]$, with the appropriate restrictions imposed at $k=-i$ and/or at $k=L+j$, where i and j are arbitrary non-negative integers.

Also shown are the expressions for w_k, $\theta = \sum_{k=0}^{L} w_k$, and $M = \sup_{(k)}.\{w_k\}$ (the maximum "height" of the lag distribution), for the special case where i=j=0, and P=N.

These latter conditions are relevant for the analysis in Chapter VIII, where the shapes depicted in Figures A.3,

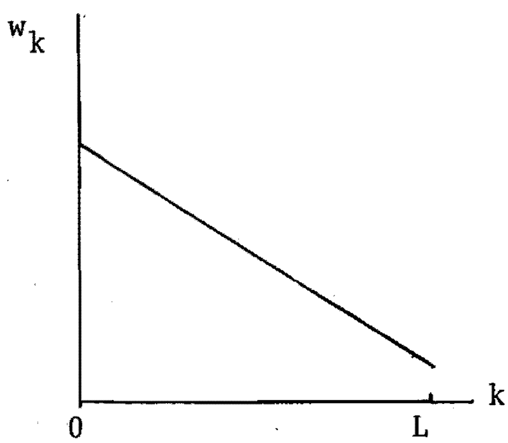


Figure A.1: P=1

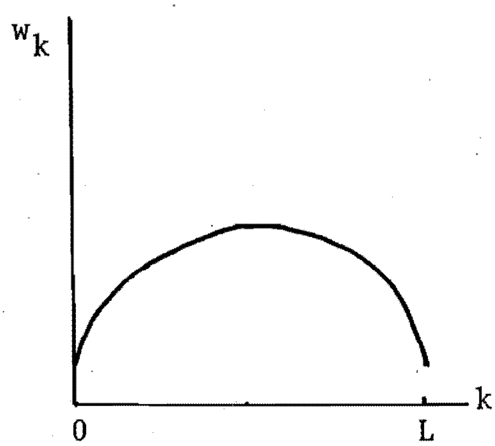


Figure A.2: $P \geq 2$

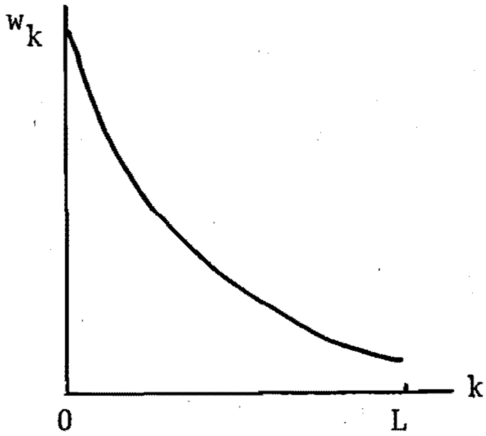


Figure A.3: $P \geq 2$

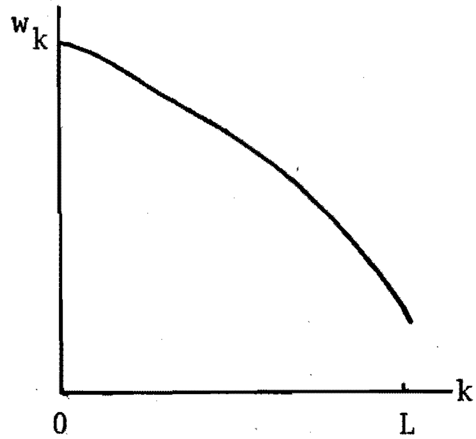


Figure A.4: $P \geq 2$

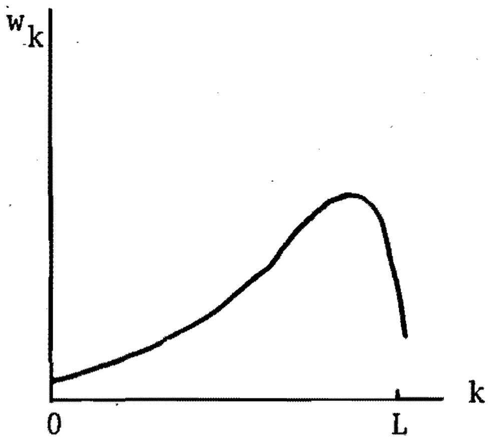


Figure A.5: $P \geq 3$

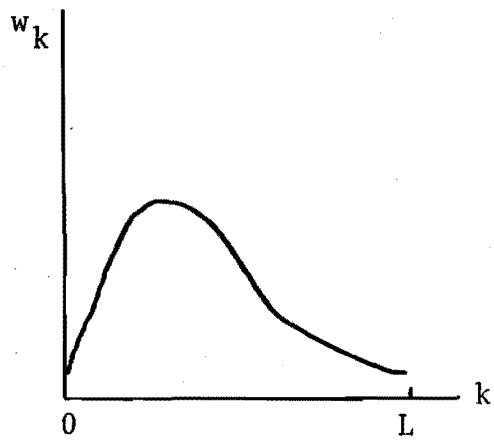


Figure A.6: $P \geq 3$

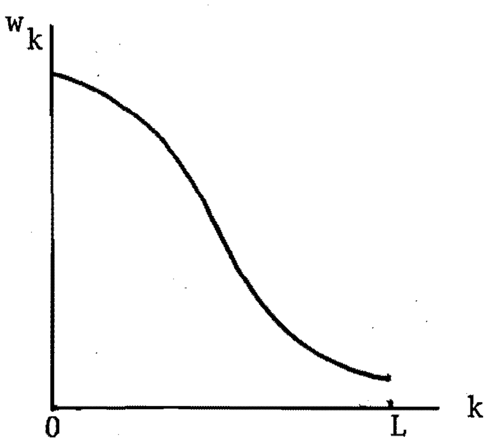


Figure A.7: $P \geq 3$

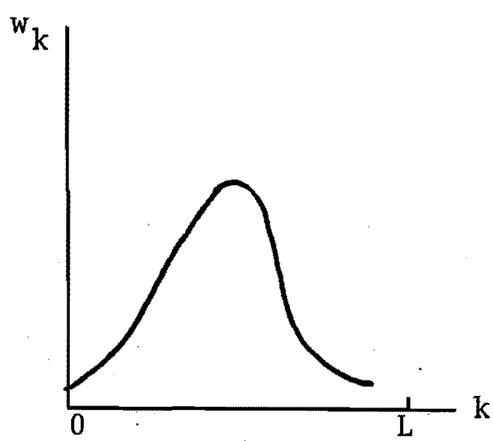


Figure A.8: $P \geq 4$

A.4 and A.7 are utilized. Of course, the condition P=N is essential to the Bayesian analysis discussed in Chapter VII. The following restrictions are imposed to attain the above shapes:

Figure A.1: $w_{L+j} = 0.$

Figure A.2: $w_{-i} = w_{L+j} = 0.$

Figure A.3: $w_{L+j} = (dw_k/dk)|_{L+j} = 0.$

Figure A.4: $w_{L+j} = (dw_k/dk)|_{-i} = 0.$

Figure A.5: $w_{-i} = w_{L+j} = (dw_k/dk)|_{-i} = 0.$

Figure A.6: $w_{-i} = w_{L+j} = (dw_k/dk)|_{L+j} = 0.$

Figure A.7: $w_{L+j} = (dw_k/dk)|_{-i} = (dw_k/dk)|_{L+j} = 0.$

Figure A.8: $w_{-i} = w_{L+j} = (dw_k/dk)|_{-i} = (dw_k/dk)|_{L+j} = 0.$

The associated expressions for w_k, θ and M when i=j=0 and P=N are:

Figure A.1: $w_k = \alpha_1(k-L)$; $\theta = -\alpha_1 L(L+1)/2$; $M = -\alpha_1 L$.

Figure A.2: $w_k = \alpha_2 k(k-L)$; $\theta = \alpha_2 L(L+1)(1-L)/6$;
$M = -\alpha_2 L^2/4$.

Figure A.3: $w_k = \alpha_2(k-L)^2$; $\theta = \alpha_2 L(L+1)(2L+1)/6$;
$M = \alpha_2 L^2$.

Figure A.4: $w_k = \alpha_2(k-L)(k+L)$; $\theta = -\alpha_2 L(L+1)(4L-1)/6$;
$M = -\alpha_2 L^2$.

<u>Figure A.5</u>: $w_k = \alpha_3 k^2(k-L)$; $\theta = -\alpha_3 L^2(L-1)(L+1)/12$;
$M = -4\alpha_3 L^3/27$.

<u>Figure A.6</u>: $w_k = \alpha_3 k(k-L)^2$; $\theta = \alpha_3 L^2(L-1)(L+1)/12$;
$M = 4\alpha_3 L^3/27$.

<u>Figure A.7</u>: $w_k = \alpha_3(L^3 - 3Lk^2 + 2k^3)/2$; $\theta = \alpha_3 L^3(L+1)/4$;
$M = \alpha_3 L^3/2$.

<u>Figure A.8</u>: $w_k = \alpha_4 k^2(k-L)^2$; $\theta = \alpha_4 L(L^2+1)(L-1)(L+1)/30$
$M = \alpha_4 L^4/16$.

Theorems A.15 and A.16 are used in deriving the above expressions for θ.

APPENDIX IV

RAW DATA FOR CHAPTER VIII

The symbols used here, and the data sources are described in Section II of Chapter VIII.

		Y	I	D
1960	(1)	123.724	122.527	-1.000
	(2)	126.731	128.982	-1.000
	(3)	143.784	144.534	-1.000
	(4)	163.369	152.792	-2.000
1961	(1)	165.062	157.282	-3.000
	(2)	153.906	172.548	-3.000
	(3)	149.560	162.376	-3.000
	(4)	125.219	125.909	-3.000
1962	(1)	114.556	116.606	-3.000
	(2)	127.344	119.280	-2.000
	(3)	126.450	151.776	-2.000
	(4)	133.978	136.156	-2.000
1963	(1)	133.872	136.945	-1.000
	(2)	151.380	144.708	-1.000
	(3)	157.430	177.092	-1.000
	(4)	166.592	163.346	-1.000

		Y	I	D
1964	(1)	152.212	159.850	-1.000
	(2)	168.486	156.098	-1.000
	(3)	164.934	175.790	-1.000
	(4)	167.410	168.982	-1.000
1965	(1)	156.130	157.388	-1.000
	(2)	172.982	165.452	-2.000
	(3)	191.040	202.204	-2.000
	(4)	196.008	186.092	-2.000
1966	(1)	182.374	185.984	-2.000
	(2)	184.476	183.204	-2.000
	(3)	179.664	183.163	-2.000
	(4)	190.428	186.176	-2.000
1967	(1)	167.746	185.306	-2.000
	(2)	172.099	181.559	-3.000
	(3)	167.205	162.090	-3.000
	(4)	148.461	127.800	-2.000
1968	(1)	144.024	173.579	-2.000
	(2)	174.592	165.673	-1.000
	(3)	193.610	208.320	-1.000
	(4)	196.084	207.382	-2.000
1969	(1)	180.391	191.140	-2.000
	(2)	201.487	200.094	-2.000
	(3)	219.248	250.014	-2.000
	(4)	214.717	221.590	-2.000

		Y	I	D
1970	(1)	212.232	215.130	-2.000
	(2)	242.301	265.687	-2.000
	(3)	264.151	278.378	-2.000
	(4)	259.416	286.050	-2.000
1971	(1)	243.682	264.554	-2.000
	(2)	258.254	272.093	-2.000
	(3)	277.419	308.094	-2.000
	(4)	272.400	284.531	-1.000
1972	(1)	245.556	259.412	-1.000
	(2)	277.196	304.015	-1.000
	(3)	279.490	300.400	-1.000
	(4)	299.396	344.100	-1.000

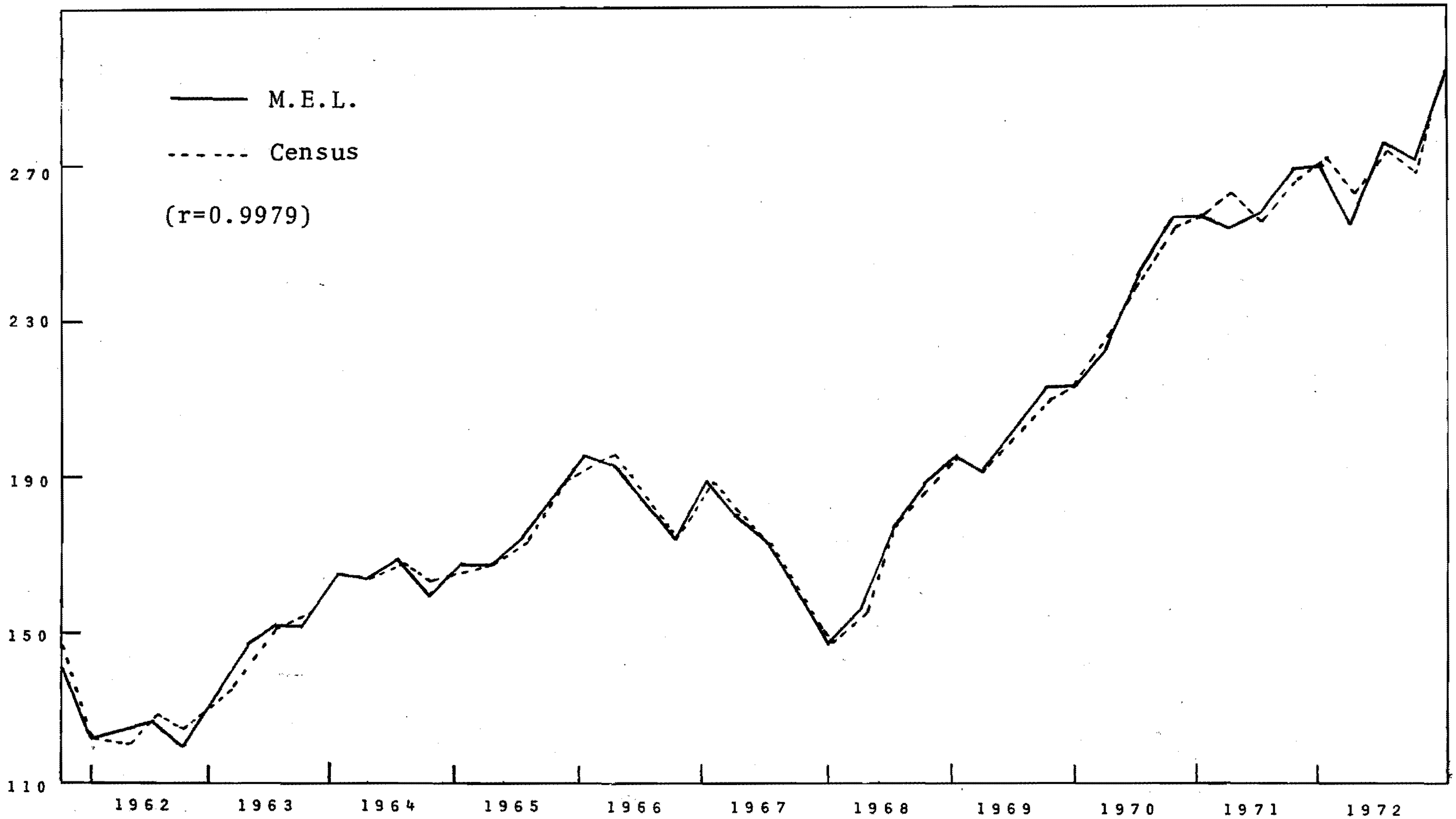


Figure A.9 : Seasonally Adjusted Current Payments.

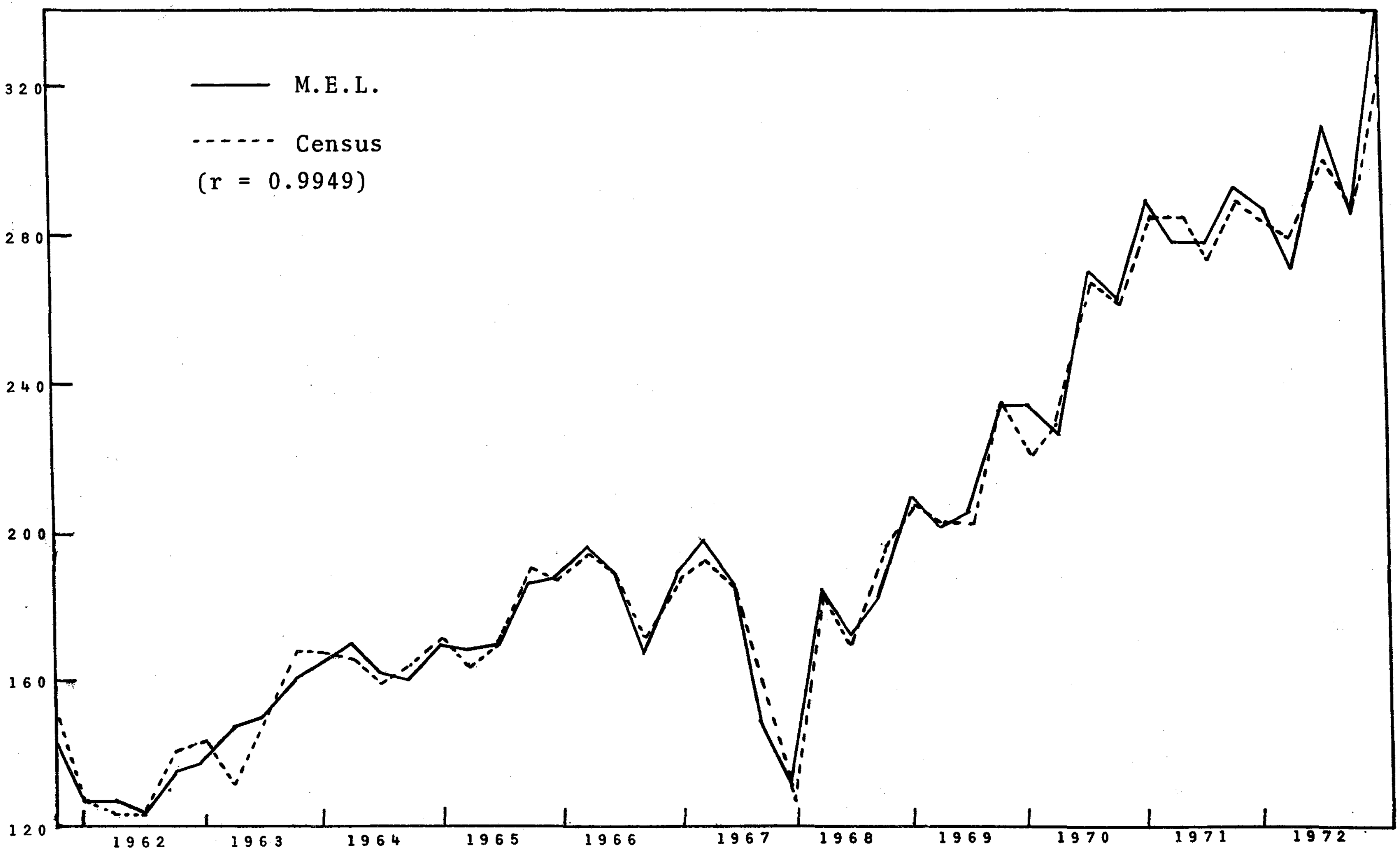


Figure A.10: Seasonally Adjusted c.i.f. Imports.